

OM of: US-09-554-547-16 to: GenEmbl:* out_format: pfs
 Date: Feb 11, 2002 12:24 PM

About: Results were produced by the Gencore software, version 4.5,
 Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:

-MODE=framer-p2n.model -DEV=xlp
 -Q=/cgn2.1/USPTO.spool/US09554547/runat.11022002.113852.17697/app_query.fasta_1.214
 -DB=GenEmbl -QFMT=lastap -SUFFIX=rg -GAP=12.000 -GAPEXT=4.000
 -MINMATCH=0.100 -LOOPEXT=0.000 -LOOPEXT=0.500 -OGAP=6.000
 -OGAPEXT=0.050 -XGAP=10.000 -XGAPEXT=0.500 -FGAP=6.000
 -FGAPEXT=0.000 -YGAPEXT=10.000 -YGAPEXT=0.500 -DELPE=6.000
 -DELPEXT=7.000 -START=1 -MATRIX=blonsum62 -TRANS=humand0.cdi
 -LIST=45 -DOCALLIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0
 -ALIGN=15 -MODE=LOCAL -OUTFMT=pis -NORM=ext -MINLEN=0
 -MAXLEN=200000000 -USER=US09554547_@CGN1_1-5156 -NCPU=6
 -ICPU=3 -LONGLOG -NO_XLPHY -WAIT -THREADS=1

Search information block:

Query: US-09-554-547-16
 Query length: 154
 Database: GenEmbl:*
 Database sequences: 1472140
 Database length: 341344837
 Search time (sec): 1273.930000

/score_list:

Sequence	Start	Orig	ZScore	Escore	Len	Documentation
gb_pat:AX003899	+	851.00	1064.17	5.3e-51	656	AX003899 Sequence 15 from Patent
gb_htg:AC003507	+	220.00	249.43	1.3e-05	149040	AP003507 Oryza sativa chrom
gb_in:PROEG91A	+	219.00	274.75	5.0e-07	1940	M71250 Drosophila melanogaster
gb_htg:AC017979	+	215.00	242.52	6.7e-06	23547	AC017979 Drosophila melanogaster
gb_in:AC009461	+	215.00	242.58	3.2e-05	171612	AC009461 Drosophila melanogaster
gb_in:AC009742	+	215.00	242.15	3.3e-05	178100	AC009742 Drosophila melanogaster
gb_in:AE003718	+	215.00	240.00	4.3e-05	253176	AE003718 Drosophila melanogaster
gb_in:AF345291	+	213.50	275.06	4.8e-07	605	AF345291 Mus musculus keratin-4
gb_htg:AP003507	+	212.00	229.52	4.6e-05	149040	AP003507 Oryza sativa chrom
gb_in:PROEG91B	+	210.50	271.48	7.6e-07	592	M71251 Drosophila melanogaster
gb_pl:AB009052	+	207.50	237.31	6.1e-05	85992	AB009052 Arabidopsis thaliana
gb_htg:AC017979	+	199.50	235.33	7.8e-05	23547	AC017979 Drosophila melanogaster
gb_in:AC009461	+	199.50	223.18	0.0004	171612	AC009461 Drosophila melanogaster
gb_in:AC009742	+	199.50	222.95	0.0004	178100	AC009742 Drosophila melanogaster
gb_in:AE003718	+	199.50	220.80	0.0005	253176	AE003718 Drosophila melanogaster
gb_gsts:CN506108	+	198.50	253.39	7.7e-06	1003	AL400761 T7 end of clone AS0A
gb_pr:HSATF168	+	198.50	248.79	1.4e-05	2126	X98893 H.sapiens mRNA for TBP
gb_pr:HSU5134	+	198.50	248.72	1.4e-05	2153	U01334 Human putative RNA bind
gb_pr:AB01005510	+	198.00	249.33	1.3e-05	1759	AB010066 Homo sapiens RBP56/h
gb_pl:AC028815	+	198.00	222.25	0.0004	147452	AC028815 Oryza sativa chrom
gb_htg:AC023247	+	198.00	221.03	0.0005	179966	AC023247 Mus musculus chrom
gb_in:AF345297	+	193.50	252.70	8.5e-06	408	AF345297 Mus musculus keratin-4
gb_in:DB9902	+	193.00	247.97	1.5e-05	798	DB9902 Mus musculus mRNA for h
gb_in:AC024790	+	192.00	223.19	0.0004	37482	AC024790 Caenorhabditis eleg
gb_in:AL451075	+	192.00	212.60	0.0011	153937	AL451075 Human DNA sequence
gb_htg:AL160286	+	192.00	212.60	0.0014	211997	AL160286 Homo sapiens chrom
gb_in:AC017920	+	191.50	230.18	0.0002	10804	AC017920 Drosophila melanogaster
gb_in:AC007723	+	191.50	213.66	0.0013	160972	AC007723 Drosophila melanogaster
gb_in:AC007646	+	191.50	213.00	0.0014	179454	AC007646 Drosophila melanogaster
gb_in:AC003701	+	191.50	212.08	0.0015	208418	AC003701 Drosophila melanogaster
gb_pr:HSK8709	+	191.00	239.05	4.9e-05	2287	X29074 H.sapiens mRNA for cyto
gb_in:AB055661	+	190.50	246.67	1.8e-05	595	AB055661 Bombyx mori GCPI mRNA
gb_pl:SCN03	+	190.50	240.61	4.0e-05	1603	X66019 S.cerevisiae NOR gene
gb_pl:SCN03	+	190.50	238.61	5.2e-05	2225	X70951 S.cerevisiae MTS1 gene
gb_pr:SE69510	+	190.50	238.27	5.4e-05	2352	S69510 Cycloketatin 9-type I cy
gb_pl:YGCNPL3A	+	190.50	234.85	8.3e-05	4111	M86731 S.cerevisiae Npl3 prot
gb_pr:HSK709	+	190.50	224.31	8.9e-05	4492	X75015 H.sapiens KRP9 gene fir
gb_in:CELT2086	+	190.50	227.24	0.0002	14262	AF000193 Caenorhabditis eleg
gb_pl:SCD9461	+	190.50	216.81	0.0008	78500	U33007 Saccharomyces cerevis
gb_htg:AC006880	+	190.50	209.87	0.0021	244239	AC006880 Caenorhabditis eleg
gb_in:AC006896	-	190.50	208.64	0.0024	298960	AC006896 Caenorhabditis eleg

gb_htg:AC013185 - 189.50 218.08 0.0007 5218 I AC013185 Drosophila melano
 gb_htg:AC016445 + 189.50 214.57 0.0011 92569 I AC016445 Drosophila melano
 gb_in:AC010211 - 189.50 210.57 0.0019 177857 I AC010211 Drosophila melano
 gb_in:AE003600 - 189.50 207.55 0.0028 291545 I AE003600 Drosophila melano

seq_name: gb_pat:AX003899

seq_documentation_block:
 LOCUS AX003899 656 bp DNA PAT 24-AUG-2000
 DEFINITION Sequence 15 from Patent WO924567.
 ACCESSION AX003899
 VERSION AX003899.1 GI:9927600

KEYWORDS
 SOURCE
 ORGANISM
 Rhinoceros appendiculatus.
 Rhinoceros appendiculatus.
 Eukaryota; Metazoa; Arthropoda; Chelicerata; Arachnida; Acari;
 Parasitiformes; Ixodida; Ixodidae; Rhinoceros.

REFERENCE
 1 (bases 1 to 656)
 Nuttall, P.A. and Paesen, G.C.
 Tissue cement proteins from rhinoceros appendiculatus

JOURNAL
 Patent: WO 924567-A 15-MAY-1999;
 NAT ENVIRONMENTAL RESEARCH COU (GB); NUTTALL PATRICIA ANN (GB)

FEATURES
 source
 1. 656
 /organism="Rhinoceros appendiculatus"
 /db_xref="taxon:34631"

BASE COUNT 160 a 158 c 191 g 147 t
 ORIGIN

alignment_scores:
 Quality: 851.00 Length: 154
 Ratio: 5.526 Gaps: 0
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
 US-09-554-547-16 x AX003899 ..

Align seg 1/1 to: AX003899 from: 1 to: 656

1 MetLysAlaPhePheValLeuSerLeuSerThrAlaAlaThrAs 17
 |||||
 47 ATGAAAGCTTTCGTCCTTCCCTTCAACCCGCGCTGACGAA 96
 |||||
 17 nAlaAlaAGAlaGlyArgLeuGlySerAspLeuSerPheGlyArg 34
 |||||
 97 TCACGACAGGCGTGGCTGCTTGGAGCAGCCTGATACATTTGGAGG 146
 |||||
 34 AlHisGlyAsnLeuTyrrAlaGlyTleGluArgAlaGlyProArgGlyTyr 50
 |||||
 147 TACACGCTAACCTATATGCGCGCATGGAAGAGCTGCGTGGATAC 196
 |||||
 51 ProGlyLeuThrAlaSerTleGlyGlyGluValGlyAlaArgLeuGly 67
 |||||
 197 CCAGGCGCTTACCGCATCGATTGGAGCGGAGTGGCGTGGCGTGG 246
 |||||
 67 yrrAlaGlyValGlyValSerSerTyrrGlyTyrGlyTyrProSerTyr 84
 |||||
 247 TCGTGGCGGTGGAGTGGAGTGGAGTGGAGTGGAGTGGAGTGGAG 296
 |||||
 84 TyrrProTyrrGlyGlyTyrGlyGlyTyrGlyGlyTyrGlyGlyTyr 100
 |||||
 297 GCGATCCGATGATGATGATGATGATGATGATGATGATGATGATGAT 346
 |||||
 101 GlyTyrrAspGlyGlyPheGlySerAlaTyrrGlyTyrProGlyTyr 117
 |||||
 347 GATATGATGATGATGATGATGATGATGATGATGATGATGATGAT 396
 |||||
 117 TgTyrrTyrrTyrrProSerGlyTyrGlyGlyGlyTyrGlyGlyTyr 134
 |||||
 397 TGGCTACTACTACTACTACTACTACTACTACTACTACTACTACTAG 446
 |||||
 134 TyGlySerTyrrGlyGlySerTyrrThrTyrrProAsnValArgAlaSerAla 150

[illegible]

```

50 yrProGlyLeuThrAlaSerIleGlyGluValGlnAlaArgLeuGly 66
81339 AT.....GGC 81335
67 GtAATgAGlAglyAlglyAlSerSerTyrGlYTyrrProSerTr 83
||||| ||| |||||
81334 GGCGGGATgATGCtgAGGA.....GGATATGCTCATcTCtGGGTa 81297
83 pGIyTYrPrOTyrGlYLYTyrrGlYGLtyrrGlyTYrTYrGLyL.... 98
:||| ||||| |||||
81296 TGcTGCTGCgtATAgctgcGAtATGccCCcCCTGcATATAgCGCGGcCT 81247
99 ..TyrGlyLyTYrAspIngIyPhGeLySerAlTyrrGlyLyTyrrPro 114
||||| ||||| |||||
81246 ACGGAGGTGcATACGCCCAgATATGcCTGtcGGAT...GCTCATcCT 81200
81199 GGTcACAcGTGGA.....GGATATGCTGCgtTACCGAGG 81165
131 ySeTYrTYrGlyLySerTYrGlyLySerTYrThrTYr 143
||||| ||||| |||||
81164 TGcTTATGCTAgTAGATACGCTGtGGGgGTGcCTAC 81128
seq_name: gb_in:DROEDG91A
seq_documentation_block:
LOCUS DROEDG91A 1940 bp DNA INV 26-APR-1993
DEFINITION Drosophila melanogaster EDG-91 gene, complete cds.
ACCESSION M71250
VERSION M71250.1 GI:157326
KEYWORDS cuticle protein.
SOURCE Drosophila melanogaster DNA.
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 1940)
Apple,R.T. and Fristrom,J.W.
20-hydroxyecdysone is required for, and negatively regulates,
transcription of Drosophila pupal cuticle protein genes
JOURNAL Dev. Biol. 146, 569-582 (1991)
MEDLINE 91323677
FEATURES
source Location/Qualifiers
1..1940
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
1053..1059
/gene="EDG-91"
join(1118..1141,1205..1660)
/gene="EDG-91"
join(1118..1141,1205..1660)
/gene="EDG-91"
/note="cuticle protein"
/codon_start=1
/protein_id="AA28502.1"
/db_xref="GI:157327"
/translation="MALVRVSCMLALLIAGQCAAPRYKTBGRLLGLGGFGSGVCI
FAGHVGSGGLYSGFGGGYPGGVASGYPGGAGGYSGNVGGSGGFGGYYPGGGYSG
SGHPRHNGGYPCGGGSYHNQGGSYGSHYSOSGYSGNGYGGGGYSGGNGFFGK."
exon /gene="EDG-91"
/number=1
1205..1738
/gene="EDG-91"
/number=2
1723..1729
/gene="EDG-91"
1739
/gene="EDG-91"
polyA_site
BASE COUNT 516 a 365 c 468 g 591 t
ORIGIN
```

```

Alignment_scores:      Quality:   219.00      Length:    172
                       Ratio:     2.235       Gaps:      11
Percent Similarity:   56.977        Percent Identity: 38.372

alignment_block:
US-09-554-547-16 x DROEDG91A ..

Align seg 1/1 to: DROEDG91A from: 1 to: 1940

7 LeuSerLeuLeuSerThrAlaAlaLeuThrAsnAlaAlaArgAlaGly 23
:::|||||: : : : : : : : : : : ||||| : : : : :
1118 ATGGCTGTGGTTCGCCGTGAAGTTCTGTAAAGTCGCGCTCATATTCCCTCC 1167

23 gluGlySer.....AspLeuAspIrrPheGlyArgValHisGlyA 37
:::|||| | ||||| ||||| : : : : :
1168 GATGGGAGCAGCATCAATGCATTTGGTTACCTTGCGAGATGGGCCCTTT 1217

37 snLeuYrAlaGlyIleGluArGAlaGlyProArgGlyYrProGlyLeu 53
||| ||||| : : : : : ||||| : : : : :
1218 TGTGATTTCGCCGTCAAGTCAAGCGCGCCCGGTGAAGACCAGAAGTCCC 1267

54 ThrAlaSer...IleGlyGlyIuValGAlaAlaArgLeuGlyIyArG 69
||| :: : : : ||||| ||||| : : : : :
1268 ACCTTGGCCCTTCTGGCGCGGTGATTTGGTGGCAGTGTAGACTTAGTGC 1317

69 acGlyValIGlyValSerSer.....TyrgLyTYrGlyYT 80
|||||: : : : : : : : : : : ||| |||||
1318 CGGCAATCGAGATGGCTGTGGCCCTGTATAGCGGTTTTGGAGGCGGTGCT 1367

80 YrPrO.....SerTrpGlyTYrProTYrGlyTYr...GlyGly 92
||||| :: : : ||||| ||||| |||||
1368 ATCTCTGCTGGCTATGCGAGTGCATACCA...GCTGATATGTGTGGTGGC 1414

93 TyrgLyGlyTYrGlyIyTYrtyrglyGlyTYrasPngInglyPheGlySerAl 109
|||||: : : : : ||||| ||||| : : : : :
1415 TACTCAGCCTATAAACCGCTACGAGGC.....AGTGAATTCGGAGCTGG 1458

109 nHyR.....GlyGlyTYr.....ProGlyTYrT 117
||||| ||||| ||||| |||||
1459 GTACTATTCAGAGGAGGATTACTCGGCTTTGGACACAGCGCCGATTATTC 1508

117 yrcGlyTYrTYrTYrProserGly.....TyrgLyGlyGly 128
:::| | | | | : : : : : ||||| : : : : :
1509 ACGGAGAGATACATATCCGGGCGGTGATCTGACACAATACAGGGCGGATCT 1558

129 TyrgLyGlySerTYr.....Gl 134
||||| ||||| ||||| |||||
1559 TATGGCGGCCACTATAGTCAGTCACAGTACTCGAATGATATTATTCGAGG 1608

134 yglSerTYrgLyGly 139
|||||: : : : : ||||| |||||
1609 TGCTGGCTATGCGAGC 1624

seq_name: gb_hcg:AC017979

seq_documentation_block:
LOCUS AC017979 23547 bp DNA HTG 09-DEC-1999
DEFINITION Drosophila melanogaster. *** SEQUENCING IN PROGRESS ***, In ordered pieces.
ACCESSION AC017979
VERSION AC017979.1 GI:5553211
KEYWORDS HTG: HTGS.PHASE2.
SOURCE fruit fly.
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 23547)
AUTHORS Adams,M. and Venter,J.C.
TITLE Direct Submission
JOURNAL Submitted (09-DEC-1999) Celera Genomics, 45 West Gude Drive,
```

```

COMMENT      Rockville, MD, USA
              This sequence was identified as CDW:10212867 by the submitter
              For more information on this record e-mail to fly@celera.com.
              * NOTE: This is a 'working draft' sequence.
              * This sequence will be replaced
              * by the finished sequence as soon as it is available and
              * the accession number will be preserved.
FEATURES
    source
        1..23547
            /organism="Drosophila melanogaster"
            /db_xref="taxon:7227"
BASE COUNT   6844 a          4896 c          4767 g          7040 t
ORIGIN
alignment_scores:
    Quality: 215.00      Length: 172
    Ratio: 2.194        Gaps: 11
    Percent Similarity: 56.977      Percent Identity: 37.791
alignment_block:
US-09-554-547-16 x AC017979  ...
Align seg 1/1 to: AC017979 from: 1 to: 23547
7 LeuSerLeuLeuSerThrAlaAlaLeuThrAsnAlaAlaArgAlaGly 23
      ::::::::::::::::::::::::::::::::::::
19905 ATGGCTGTTCGGTCGTGAGTGGTAAGCCGGCTGCTATTTCGGCTCC 19954
23 gLueGlySer.....AspleuaspThrPheGlyAlaGValLhISGLYA 37
      :::::|||||
19955 GATTCGGGATGCACATGCATATTTGGTTACCTTGCAGATGCGCCCTTT 20004
37 snLeuTyAlaGlyIleGluAlaArgAlaGlyProArgIlyTyProGlyLeu 53
      :::::|||||
20005 TGCTGATTGGCCGCTCAAGTCACAGCGCGCGCGGTGAAGACGAGATGAC 20054
54 ThrAlaSer...IleGlyGlyGluValAlaAlaArgLeuGlyAla 69
      ||| :::::|||||
20055 ACCTTGGGCCCTTCGCGCGGTGGATTTGGTGGCAGTGAAGACTTAAGTGC 20104
69 aGlyValGlyAlaSerSer.....TyArgIlyTyArgIlyT 80
      |||:::|||||
20105 CGGCATCGAGATGGGTGGTGGCTGCTATACGGCTTTCGAGAGCGGTGCT 20154
80 TyrPro.....SerTyrProTyrProTyrGlyIlyTyr..GlyIly 92
      ||||| :::::|||||
20155 ATCCGTGGGCTATGACGATGGCTACCCA...GTTGGCTATGTTGGTGGC 20201
93 TyrGlyGlyTyArgIlyTyArgIlyTyArgIlyTyAspGlnGlyPheGlySerAl 109
20202 TACTCAGGCTATTAACGGCTACGGAGGC.....AGTGGATTGGAGAGTGG 20245
109 aTyr.....GlyGlyTyProGlyTyTyr.....T 117
      ::||| |||||
20246 CTACTATCCAGAGAGAGATTACTCCGGCTTTGGACACAGCGCATCAC 20295
117 TyrGlyTyTyrTyProSerGly.....TyrGlyGlyIly 128
      ::||| |||||
20296 ACGGAGGATCTATACCGGGGGGTGATGATCAACCAATCAGGGCGCATCT 20345
129 TyrGlyGlySerTyr.....GAT 134
      |||||
20346 TATGCGCGCCACTATACACTCAGTCACTACTCAATGATATTACGAGAG 20395
134 yGlySerTyArgIly 139
      |||:::|||||
20396 TGGTGGCTATGAGAGC 20411
seq_name: gb_in:AC009461
seq_documentation_block:
LOCUS      AC009461 171612 bp DNA INV 23-MAR-2001

```

Jeegmann, C., Caldwell, M., Aulsebrook, D., Liff, P., Matteoli, B., Mos
McIntosh, T. C., Moy, M., Murphy, B., Nelson, C., Nelson, K.
Pacelle, J., Paragas, V., Park, S., Patel, S., Pfeiffer, B.,
Piouanenavong, S., Pittman, G. S., Puri, V., Richards, S.,

```

TITLE
JOURNAL
REFERENCE
AUTHORS
Zaveril,J.S., Smith,H.O., Rubin,G.M. and Venter,J.C.
Sequencing of Drosophila chromosome 3R, region 90C-90C
Unpublished
2 (bases 1 to 178100)
Celniker,S.E., Agbayani,A., Arcana,T.T., Baxter,E., Blazek,R.G.,
Butenhoff,C., Champe,M., Chavez,C., Chew,M., Ciesiolka,L.,
Doyle,C.M., Farfan,D.E., Galle,R., George,R.A., Harris,N.L.,
Hoskins,R.A., Houston,K.A., Hummstl,S.R., Karra,K., Kearney,L.,
Kim,E., Lee,B., Lewis,S., Li,P., Lomutan,M.A., Maada,P.,
Moshrefi,A.R., Moshrefi,M., Nixon,K., Pacleb,J.M., Park,S.,
Pfeiffer,B., Poon,L., Sequeira,A., Sethi,H., Snir,E.,
Svitsksys,R.R., Wan,K.H., Weinburg,T., Zhang,R., Zieran,L.L. and
Rubin,G.M.
Direct Submission
Submitted (30-AUG-1999) Drosophila Genome Center, Lawrence Berkeley
Laboratory, MS 64-121, Berkeley, CA 94720, USA
On Mar 14, 2001 this sequence version replaced gi:5922029.
Sequence submitted by:
Berkeley Drosophila Genome Project
Lawrence Berkeley National Laboratory, MS 64-121
Berkeley, CA 94720
This sequence was assembled using end sequences from a whole genome
shotgun and from subclones of this BAC and its neighboring clones.
For further information about this sequence, including its location
and relationship to other sequences, please visit our sequence
archive Web site (http://www.fruitfly.org/sequence/) or send email
to dbgap@fruitfly.berkeley.edu.
FEATURES
source
1..178100
/organism="Drosophila melanogaster"
/strain="y: cn bw sp"
/db_xref="taxon:7227"
/chromosome="3R"
/map="90C-90C"
/clone="BACR21K03 (D982)."
/clone_11b="RPC1-98 (Roswell Park Cancer Institute
Drosophila melanogaster BAC library, partial ECORI in
pBAC3.6)"
BASE COUNT
51308 a 37603 c 37904 g 51265 t
ORIGIN
alignment_scores:
Quality: 215.00 Length: 172
Ratio: 2.194 Gaps: 11
Percent Similarity: 56.977 Percent Identity: 37.791
Alignment block:
US-09-554-547-16 x AC009742/rev ..
Align seq 1/1 to reverse of: AC009742 from: 1 to: 178100
7 leusertleusesthrvalaaleuThrAsnAlaIargAlaIaglyur 23
:::|||||: :|||||: :|||||: :|||||:
45483 ATGGCTTGCGTTCGGCTGAGTGTGTGAAGTCCGGCTGCTATCCCGCTCC 45434
23 gLeuGlySer.....AspleuAsprhrpheiIatrgValInisGlyA 37
:::|||||: :|||||: :|||||: :|||||:
45433 GATTGGATGCACATGATCGATTGGTTCACCTTGCAAGATGCTGGCCCTTT 45384
37 snleuTfAlaIaglyIleGluArgAlaGlyProArgcIyTyrProGlyLeu 53
:::|||||: :|||||: :|||||: :|||||:
45383 TCCGTGATTCGCCGCGTCAGAGTCCAGCGCGCGCGGTGAAGACCGAAGCTTAC 45334
54 ThrAlaSer...IleGlyGlyGluValGlyAlaArgLeuGlyLysGly 69
:::|||||: :|||||: :|||||: :|||||:
45333 ACCTTGGCCCTTCGGCGGTGATTTGGTGGCAGTGTAGACACTTATGTC 45284
69 aGlyValcIValIserSer.....TyrGlyTyrGlyT 80
:::|||||: :|||||: :|||||: :|||||:
45283 CCGCATGGAGTGGCTGGCTGGCTGTATAGCGGTTTTCGAGAGCGGCTGCT 45234
80 yrPro.....SerTPrGlyTyrProTyrGlyGlyTyr...GlyGly 92

```

seq_name: gb_in:AE003718

	seq_documentation_block:	DNA	INX	DATE
LOCUS	AE003718	233176 bp		05-OCT-2000
DEFINITION	Drosophila melanogaster genomic scaffold 142000013386035 section 43 of 105, complete sequence.			
ACCESSION	AE003718 AE002708			
VERSION	AE003718.2 GI:10726582			
KEYWORDS	HNC.			
SOURCE	Fruit fly. <i>Drosophila melanogaster</i> Eukaryota; Neozoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Meoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.			
ORGANISM				
REFERENCE	1 (bases 1 to 253176) Adams,M.D., Celisner,S.E., Holt,R.A., Evans,C.A., Gocayne,J.D., Amanatides,P.G., Scherer,S.E., Li,P.W., Hoskins,R.A., Galie,R.F., George,R.A., Lewis,S.E., Richards,S., Ashburner,M., Henderson,S.N., Sutton,G.C., Wortman,J.R., Yeandell,M.D., Zhang,Q., Chen,L.X., Brandon,R.C., Rogers,Y.H., Blazey,R.G., Champe,M., Pfeiffer,B.D., Wan,K.H., Doyle,C., Baxter,E.G., Helt,G., Nelson,C.R., Gabor Milos,G.L., Abril,J.F., Agbayani,A., An,H.J., Andrews-Pfannkuch,C., Baldwin,D., Ballew,R.M., Basu,A., Baxendale,J., Bayraktaroglu,L., Beasley,E.M., Beeson,K.Y., Benos,P.V., Bernan,B.P., Bhandari,D., Bolshakov,S., Borokova,D., Botchan,M.R., Bouck,J.J., Broksrein,P., Brotlier,P., Burris,K.C., Busam,D.A., Butler,H., Cadieu,E., Center,A., Chandra,I., Chenry,J.M., Cawley,S., Dahlke,C., Davenport,L.B., Davies,P., de Pallos,B., Delcher,A., Deng,Z., Mays,A.D., Dew,I., Dietz,S.M., Dodson,K., Doup,L.E., Downes,M., Dugan-Rocha,S., Dunkov,B.C., Dunn,P., Durbin,K.J., Evangelista,C.C., Ferrara,C., Ferriere,S., Flaischmann,W., Foster,C., Gabrielian,A.E., Gary,N.S., Gelbart,W.M., Glasser,K., Glodok,A., Gong,F., Gorrell,J.H., Gu,Z., Guan,P., Harris,M., Harris,N.L., Harvey,D., Helms,T.J., Hernandez,J.R., Houck,J., Hostin,D., Houston,K.A., Howland,T.J., Wei,M.H., Ibegwan,C., Jalali,M., Kalush,F., Karpen,G.H., Ke-Z., Kemmison,J.A., Ketchum,K.A., Kimmel,B.E., Kodira,C.D., Kraft,C., Kravitz,S., Kulp,D., Lai,Z., Lasko,P., Lei,Y., Levitsky,A.A., Li,J., Li.Z., Liang,Y., Lin,X., Liu,X., Maltef,B., McIntosh,T.C., McLeod,M.P., McPherson,D., Mekullov,G., Mishina,N.V., Mobarry,C., Morris,J., Mosherel,A., Mount,S.M., Moy,M., Murphy,B., Murphy,L., Murzy,D.M., Nelson,D.L., Nelson,D.R., Nelson,K.A., Nixon,K., Nusskern,D.R., Pacleib,J.M., Palazzolo,M., Pittman,G.S., Pan,S., Pollard,U., Puri,V., Reese,M.G., Reinert,K., Remington,K., Saunders,R.D., Scheeler,F., Shen,H., Shue,B.C., Siden-Klamos,I., Simmons,M., Skupski,M.P., Smith,T., Spier,E., Spradling,A.C., Staplston,M., Strong,R., Sun,E., Svrtkus,R., Tector,C., Turner,R., Venter,E., Wang,A.H., Wang,X., Wang,Z.Y., Wasserman,D.A., Weinstock,G.M., Weissbach,U., Williams,S.M., Woodage,T., Wooley,K.C., Wu,D., Yang,S., Yao,O.A., Ye,J., Yen,R.F.,			

Zaveri, J.S., Zhan, M., Zhang, G., Zhao, Q., Zheng, L., Zheng, X.H.,
Zhong, F.N., Zhong, W., Zhou, X., Zhu, S., Zhu, X., Smith, H.O.,
Gibbs, R.A., Myers, E.W., Rubin, G.M. and Venter, J.C.
The genome sequence of *Drosophila melanogaster*
JOURNAL
TITLE
20196006
2 (bases 1 to 253176)
REFERENCE
AUTHORS
JOURNAL
Submitted
Direct Submission
Submitted (21-Mar-2000) Celera Genomics, 45 West Gude Drive,
Rockville, MD, USA
COMMENT
On Oct 9, 2000 this sequence version replaced gi:7300262.
FEATURES
source
1. 253176
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/chromosome="3R"
complement(join(<9015..9066,9230..9493,9803..10067,
10198..10425,10609..10857,11045..11208,11695..12010,
12723..12848,13260..13634,13863..14042,14361..14516,
14593..14901,15656..15799,15889..16144,16308..16518,
16861..17087,17202..17380,17568..17597,17972..18091,
18411..18546,18901..19046,19141..19290,19363..19494,
21408..21612,21940..22040,22140..22267,23854..24077,
24436..24553,25877..25997,26180..26329,26586..26747,
27102..27144))
/gene="CG12536"
/product="CT33960"
/db_xref="FLYBASE:FBan0012536"
/db_xref="FLYBASE:FBgn0038513"
complement(<9015..>27144)
/gene="CG12536"
/map="90B1-90B2"
/db_xref="FLYBASE:FBan0012536"
/db_xref="FLYBASE:FBgn0038513"
complement(join(9015..9066,9230..9493,9803..10067,
10198..10425,10609..10857,11045..11208,11695..12010,
12723..12848,13260..13634,13863..14042,14361..14516,
14593..14901,15656..15799,15889..16144,16308..16518,
16861..17087,17202..17380,17568..17597,17972..18091,
18411..18546,18901..19046,19141..19290,19363..19494,
21408..21612,21940..22040,22140..22267,23854..24077,
24436..24553,25877..25997,26180..26329,26586..26747,
27102..27144))
/note="CG12536 gene product"
/codon_start=1
/db_xref="FLYBASE:FBan0012536"
/db_xref="FLYBASE:FBgn0038513"
/protein_id="AAF5425.1"
/db_xref="GI:7300263"
/translation="MRSRRRSGSTAVLRORPHVHENTEVYLGNAIKALPEY
VRPYRVASWHRGEITLPDLSVAGRVVYLAASGDLVRSVRSEDEGKPSCLVNT
LNGEKORDAVMLQVELSKNLAPPTTKPMEHVERGNDVHLPCNQGNEPLFTW
YRVSALALYIPSSORVLTSLRTLLIKNADKRDAGKWIQASNGFGORIRLSVN
SYSVHILPQVQIVNSGCTANCTTGSIDAIDMLNGKPLQANALITGGRNIR
LSKSLVONQGRDGRGVYOCLEVNORASQAQMAELKLDVPELITFIETONRPG
LISLKCSAGSSPPQFAWLLDSOPTMDYSLHPRALIQFVMSGDVLSHLNLSVRPD
DGGLKCYASMSKGSVOHSAKLANYPYVAIRKIRKAVAGEDITVHCPFGYVEOI
RWEKHAQELTTSNHELSVADGGOLVKNRPGDGIYTCIVRSRGEARDEMOL
NVNPSVLEPFRPKNLOEGRAQITCAVSSDMPIYFSMKDDSSIPSSLOITEKE
EYSLVLEKDISARHSGKYCYASMAAKANVYTAELQYVAPRMRYEMPMADIMGNT
ISINCEAGYPIPTTWKGGGSKDKPLKPAPEFASANISRRNDPTLCHAK
GDEPTTIGWTONNGRIDLNFRPSTAEKTEKGERPDPSHLETFEFGSRPVKLSMR
PFQNSPVLSLVQYQALKTQSHSGSLAAGDMGHVINSLSPTSISRSYDSLR
SAVAGLTPTATFLIRMQALNIEKSAATEAIVLKTQEAPEARSNVOVQGESE
IVTQVAPRESMNGELIGYVNCSEKONIFISVNSLSKSTISVSGWATKATIRGL
KRYSRVATIRAMNSFGSPMSAIFGTAGVEAONVNCALSSQSLKATLRL
PLOGHGIIIGYKILIRPIVHIDPEAFITKRSNLETYHLTHKASNYSTRVLAAT
ATGGGLASHPLFCQTDVDPDPAIRAKAALTADSILSLMTPKRNCTISHYVYSR
EAGRGAQKTHVRVDENGVPTFESRSLANQWEFVNSVSTSGESEPSTVAQAT
NTRPARIASFGVVRKAVGTGLVLECLAVNQPTRAWMLTRDRPVFSPYEVTNEG
NLKIRVSGSLSGNVTYANNLFGSDEIOYVIAMKPPSAPOIIVQYASASIRVNSD

APDGGAPLQGVTSYHTAGSWSITELLPPENMAFTISGLCNGOYITIKSAHNWGS
GVASEEIVWTKGKASQAPNANIELATVATVNLKLSQMWGSGSIHFFSTIEHPLG
IRWVVTSDIINAENRENLIFCDPLKAVQOLRSATNDAGKTEHHEFTSTPDGI
TTPSPVSPENDMLNNLINSTNPSGWFNTLLVYVITYSITLITATIRHRTG
PIABGESRTLPQDKEDHENRMOQVYASAPVKTVDKNSEMEITPYATFESYNG
RTGAPAKPTPRAVAQPLDITWQPKFTGHEGDELNATVAPLPSGSGVKSXSSW
HKORYNNTEDSTLSKSMITVAGSQAGSKKSNGRSKSSASAGSVAGSESDPIS
STEFESNMPRTVPCKSSRSSDGRAVYMPDPDSSTESNDGSPAPERRHNTPHVL
GGMAMGIGGGGGVGGGNGPAERKRSAGRSRKHGSAAOQPSNOTLERKCPGSSNS
LDSVYDETASIASIAAMSGAGADLSCGFPRAFGHGREPDGSDDERERLAL
DLEVQWVESTGDQOLAKMDEELTSLARHEKKEDEKRELYTTHV"
complement(join(31465..31726,32418..32700,38192..38335,
39745..39767))
/gene="CG8106"
/product="CT24226"
/db_xref="FLYBASE:FBan0008106"
/db_xref="FLYBASE:FBgn0038514"
complement(<31465..>39767)
/gene="CG8106"
/map="90B1-90B3"
/db_xref="FLYBASE:FBan0008106"
/db_xref="FLYBASE:FBgn0038514"
complement(join(31503..31726,32418..32457))
/gene="CG8106"
/note="CG8106 gene product"
/codon_start=1
/db_xref="FLYBASE:FBan0008106"
/db_xref="FLYBASE:FBgn0038514"
/protein_id="AAF55426.1"
/db_xref="GI:7300264"
/translation="MSLPEPCVLVLEKALLGALLAVLATPLATSLRVPTLLEP
APRLFLGNDTGAQVCTVHSGSPPLVTVLVLDGSLATVPELR"
complement(join(66536..67410,67471..67541,67834..68268,
68330..68753))
/gene="CG5823"
/product="CT18279"
/db_xref="FLYBASE:FBan0005823"
/db_xref="FLYBASE:FBgn0038515"
complement(<66536..>68753)
/gene="CG5823"
/map="90B3-90B4"
/db_xref="FLYBASE:FBan0005823"
/db_xref="FLYBASE:FBgn0038515"
complement(join(66931..67410,67471..67541,67834..68134))
/gene="CG5823"
/note="CG5823 gene product"
/codon_start=1
/db_xref="FLYBASE:FBan0005823"
/db_xref="FLYBASE:FBgn0038515"
/protein_id="AAF55427.2"
/db_xref="GI:10726583"
/translation="MSSSTSGGKKOPAVSRMKODYRKLKRDPLPYTAPPLNNIL
EMHVCVGPEDSPYGGYHGTLLPREFPEKPSVITLTNGRKTKTLCISDR
HPTDWNLPWCVTGLTGLISLPLSTPLGISSSNDKQMFQAKSLAFNLRNIFCE
LPELIVEIKIQQTQAAAVDPAAGVRSRSLANGASLADGSLAVAGLANPL
NAVAIDCGAAGQADLASGGAAYONSARNSYLWMQSVYSLVITICPAIPALIVNY
IKLNQOE"
join(69939..70062,70244..70909,70982..>71058)
/gene="CG5840"
/product="CT18319"
/db_xref="FLYBASE:FBan0005840"
/db_xref="FLYBASE:FBgn0038516"
complement(<71058..>71058)
/gene="CG5840"
/map="90B3-90B4"
/db_xref="FLYBASE:FBan0005840"
/db_xref="FLYBASE:FBgn0038516"
join(69984..70062,70244..70909,70982..71058)
/gene="CG5840"
/note="CG5840 gene product"
/codon_start=1
/db_xref="FLYBASE:FBan0005840"
/db_xref="FLYBASE:FBgn0038516"


```

136 rTyrglygTySerTyr 141
      :|||||:
      356 ATATGCTCTGCTTC 371
seq_name: gb_hcg:AP003507

seq_documentation_block:
LOCUS      AP003507      149040 bp      DNA      HTG      12-APR-2001
DEFINITION Oryza sativa chromosome 6 clone P0468A12, *** SEQUENCING IN
PROGRESS ***, in ordered pieces.
ACCESSION  AP003507
VERSION     AP003507.1  GI:13603466
KEYWORDS   HTG; HTGS_PHASE2.
SOURCE      Oryza sativa (cultivar:Nipponbare) DNA, clone:P0468A12.
ORGANISM    Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
              Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
              Ehrhartoideae; Oryzaceae; Oryza.
REFERENCE   1 (sites)
AUTHORS     Sasaki,T., Matsumoto,T. and Yamamoto,K.
TITLE        Oryza sativa nipponbare(GA3) genomic DNA, chromosome 6, PAC
              clone:P0468A12
JOURNALT    Published Only in Database (2001) In press
REFERENCE   2 (bases 1 to 149040)
AUTHORS      Sasaki,T., Matsumoto,T. and Yamamoto,K.
TITLE        Direct Submission
JOURNAL      Submitted (11-APR-2001) Takuji Sasaki, National Institute of
              Agricultural Resources, Rice Genome Research Program, Kannondai
              2-1-2, Tsukuba, Ibaraki 305-8602, Japan
              (E-mail:tsasaki@ab.affrc.go.jp, URL:http://rgp.dna.affrc.go.jp/,
              Tel:81-298-38-7441, Fax:81-298-38-7468)
NOTE        NOTE: It currently consists of 1 contigs. Gaps between the contigs
              are represented as runs of N. The order of the pieces is believed
              to be correct as given, however the sizes of the gaps between them
              are based on estimates that have provided by the submitter. This
              sequence will be replaced by the finished sequence as soon as it is
              available and the accession number will be preserved.
              * NOTE: This is a 'working draft' sequence.
              * This sequence will be replaced
              * by the finished sequence as soon as it is available and
              * the accession number will be preserved.
FEATURES
  source
    1..149040
       /organism="Oryza sativa"
       /cultivar="Nipponbare"
       /db_xref="taxon:4530"
       /chromosome="6"
       /clone="P0468A12"
BASE COUNT  41830 a 33432 c 32383 g 40945 t 450 others
ORIGIN
alignment_scores:
  Quality: 212.00      Length: 103
  Ratio: 3.212         Gaps: 5
  Percent Similarity: 64.078      Percent Identity: 49.515
alignment_block:
US-09-554-547-16 x AP003507 ..
Align seg 1/1 to: AP003507 from: 1 to: 149040
47 ProArGlyTYrProGlyLeuThrAlaSerIleGlyGluValGlyAl 63
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
109383 CCTACGGAGGGCCAGCGGTTGAGATCAAAAGTGGGAGCT...GGATA 109429
   63 aArGleuGlyArgAlaGlyValGlyValSerSerTyrgly.....T 78
   ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
109430 CTAACCTGGTGGAGGATATGATATGT...GTTGGTACGGTGGAGAT 109476
   78 yrglyTYrProSerTyrglyTYrProTyrglyGlyTyrglyTYrgly 94
   ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

109477 ATGCTCCTCGGTACGGTGTGGGTACGGTGAGCATATGCTATCCT 109526
      95 GlyTyrglyGly.....TyrglyGlyTyraSpGlnGlypheGlyserA 109
      :|||||:
      109527 GGCTATGCTGTGGTGCATACGGCGGTGATATGACGTGGGTATGCTGCCG 109576
109      atTyrglyGlyTyrProGlyTYrTyrglyTYrTyProSerGlyTyrg 126
      :|||||:
109577 GTACGGTGTGGTGGCGGTGAGCATATGCGGTGATACGAGGTGATACG 109626
126 lyGlyGlyTy.....GlyGlySerTyrglyGlySerTyrglyGly 140
      :|||||:
109627 GAGGTGATACGGTGTGGTGGGTTACGGTGAGATATGTTGGT 109676
141 TyrThrTyr 143
      |||
109677 GGTGGTAC 109685
seq_name: gb_in:DROEDG91B

seq_documentation_block:
LOCUS      DROEDG91B      592 bp      mRNA      INV      26-APR-1993
DEFINITION Drosophila melanogaster EDG-1 cuticle protein mRNA, complete cds.
ACCESSION  M71251
VERSION     M71251.1  GI:157328
KEYWORDS   cuticle protein.
SOURCE      Drosophila melanogaster cDNA to mRNA.
ORGANISM    Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
              Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
              Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1 (bases 1 to 592)
AUTHORS     Apple,R.T. and Fristrom,J.W.
TITLE        20-hydroxyecdysone is required for, and negatively regulates,
              transcription of Drosophila pupal cuticle protein genes
              Dev. Biol. 146, 569-582 (1991)
JOURNALT    MEDLINE
MEDLINE     9132677
FEATURES
  source
    1..592
       /organism="Drosophila melanogaster"
       /db_xref="taxon:7227"
       /gene="EDG-91"
       /gene="EDG-91"
       /gene="EDG-91"
       /note="cuticle protein"
       /codon_start=1
       /protein_id="AA028503.1"
       /db_xref="GI:157329"
       /translation="MALVRRVSCMLALLLIAGQGAAEVKTEGRTGLDGGFGGSGVL
              SAGIVGGLYSGFGGCGYPGYASGYPGAGGYSGVNGYSGSGFGGYYPGGYSGL
              FGHRPHYGGYVPGCGSYVHNGCGSYGSHYSQYSNGLYGGGCGGNGPFGR"
              577..583
polyA_signal
BASE COUNT  126 a 120 c 198 g 148 t
ORIGIN
alignment_scores:
  Quality: 210.50      Length: 138
  Ratio: 2.699         Gaps: 10
  Percent Similarity: 56.522      Percent Identity: 42.754
alignment_block:
US-09-554-547-16 x DROEDG91B ..
Align seg 1/1 to: DROEDG91B from: 1 to: 592
38 leuTyraGlyYlIleGluArgAlaGlyProArgGlyTyProGlyLeuTh 54
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
74 CTGATTGCCGTCAGAGTCAGCGCGGTGAAGACCGAAGGTCGCAC 123
54 rAlaser...IleGlyGlyGluValGlyAlaArgLeuGlyArgAlaG 70

```


[illegible]

<http://gremlini.zozi.iastate.edu/cg1-bin/sp.cgi>).
 Genes encoding tRNAs are predicted by tRNAscan-SE
 (Sean Eddy, Washington University School of Medicine, St. Louis,
<http://genome.wustl.edu/peddy/tRNAscan-SE/>).
 This sequence may not be the entire insert of this clone. It may be
 shorter because we remove overlaps between neighboring submissions.
 The 5' clone is K2116 and the 3' clone is K1816.

Location/Qualifiers

1..85392

/organism="Arabidopsis thaliana"
 /strain="Columbia"
 /db_xref="taxon:3702"
 /chromosome="5"
 /clone="MNF13"
 /clone_11b="Mtsul p1"
 join(244..368,668..708,812..975,1066..1165,1255..1424
 1513..2184)
 /note="contains similarity to RNA binding protein
 gene_id:MNF13.1"
 /codon_start=1
 /evidence="not_experimental"
 /protein_id="BAB08520.1"
 /db_xref="GI:9758076"
 /translation="MDPEQETVEFEDERHDPKSEDIEDDDKRSOPHGGGVSAGKTI
 EVGGLARTTSAEFLKAEKAGKGETTDSYIMKDRKTGQPRGFTYADSSVNDKVIDD
 NHIIIGKVEKTRITIPGMSNDPFTKRIIVGGISVDDDEKFEFFWQGLKEKHQ
 IMRHSGTRSGFGFVYESDDMDHLIAKNTELSTGVQVEIKKAPKPKNSVTGPS
 KRFDGSRNPFGGGAGCGGCGGCGGPGPYKSGGCGGGRSGYCGYGGEGRGYG
 GGGGAGGAPVGRPEALDLSGRYGGGGGYNRGYSMGGGGYCGGPDYGGSGYGP
 GGGYGGPSGGGSGGIGGCGGCGAGCGGRCGGGIDMGCGVGGGAGCGAGAG
 CGNGGSGGGGCGRGCGTGGGSGSRHPHYGR"
 join(4110..4271,4372..4445,4682..4987,5065..5148,
 join(4110..4271,4372..4445,4682..4987,5065..5148,

```
seq_documentation_block:
LOCUS      AB009052      85992 bp      DNA      27-DEC-2000
DEFINITION Arbidopsis thaliana genomic DNA, chromosome 5, p1 clone:IMN13
ACCESSION  AB009052 BA000015
VERSION    AB009052.1 GI:265028
KEYWORDS
SOURCE     Arbidopsis thaliana (strain:Columbia) DNA, clone_1lib:Mitsui p1
```

SOURCE	ORGANISM
Arabidopsis thaliana (strain:Columbia) DNA, clone:lib:Mitsui P1 Clone:MNF13.	Arabidopsis thaliana
	Eukaryota: Viridiplantae: Streptophyta: Embryophyta: Tracheophyta: Spermatophyta: Magnoliophyta: eudicotyledons: core eudicots: Rosidae: eurosids II: Brassicales: Brassicaceae: Arabidopsis. 1 (slices)
REFERENCE	
AUTHORS	Sato, S., Kaneko, T., Kotani, H., Nakamura, Y., Asamizu, E., Miyajima, N and Tabata, S.

TITLE Structural analysis of Arabidopsis thaliana chromosome 5. IV. Sequence features of the regions of 1,456,315 bp covered by nineteen physically assigned P1 and YAC clones

JOURNAL DNA Res. 5 (1), 41-54 (1996)

MEDLINE 98290546
REFERENCE 2 (bases 1 to 85992)

AUTHORS	Nakamura, Y.
TITLE	Direct Submission
JOURNAL	Submitted (27-NOV-1997) Yasukazu Nakamura, Kazusa DNA Research

COMMENTS

For the latest information on annotation of this clone, please see http://www.kausa.or.jp/kaos/cgi-bin/agg_graph.cgi?c=NNRF13. Genes with similarity to proteins in the databases are described in 'Genes' with similarity to proteins in the databases are described in 'product' or 'note' qualifiers. Genes that have no significant protein similarity are described as 'unknown protein'. The software programs used to predict genes include: Grail (Informatics Group, Oak Ridge National Laboratory, <http://compbio.cornell.gov/Grail-1.3/>),

GENSCAN (Chris Burge, MIT, <http://CCR-081.mt.edu/GENSCAN.html>)
NetGene2 (S.M. Hebsgaard, et al., CBS, Technical University of Denmark, <http://www.cbs.dtu.dk/services/NetGene2/>) and
SplicePredictor (Volker Brendel, Stanford University,

FEATURES

Location/Qualifiers

<http://gremlini.zool.iastate.edu/cgi-bin/sp.cgi>.
Genes encoding tRNAs are predicted by tRNAscan-SE
(Sean Eddy, Washington University School of Medicine, St. Louis,
<http://genome.wustl.edu/eddy/tRNAscan-SE/>).
This sequence may not be the entire insert of this clone. It may be
shorter because we remove overlaps between neighboring submissions.
The 5' clone is K21116 and the 3' clone is K1B16.

Source

CDS

```
/evidence=not_experimental  
/protein_id="BAB08520.1"
```

/db_xref="GI:9758076"
 /translation="MDPGRGVTEIEDIEDHPKSPEDIEDDDKSKQPHSGGCVSAGRI
 FVGGLARPTAEFLKTHGKAGETEDYIMDRDTPQPRGFEPTVADSSVDDVAD
 NHIIIGKVEIKRTIPROGSMNDPFTKRIIVGIGIPSSVDDEKFEFMOGELKHO
 IMPHDSRSGFGFVPESEMDYHLAKNRNITESTGVETIKAPKPNVSTYPS
 KRPFDSRNPCCGGYDDCGGGHGGGCGGPGPYRSGGCGVGGGCGGGEAGG
 CGGCGGCGVPRGCEPALDYSRYRGGGGCGVYRGCSMGGGGCGGCGCDMAYGSEYR
 CGGGRPGSGTGGGIGCGTGCGGAGCGGGRGGGGTIDMGVGGGAGCGTGAAG
 CGGCGSGSYGGGCGGCGGGGSGRHYPLG"
 join(4110..4271,4372..4445,4682..4987,5065..5148,
 5256..5559)

CDS

/note="gene_id:MF13.2"
 /codon_start=1
 /evidence=not_experimental
 /product="sucrose cleavage protein-like"
 /protein_id="BA08521.1"
 /db_xref="GI:9758077"
 /translation="WYNNINASTSYARAHVLYKTPPEALSHVEEGLPQRAITLL
 KDRSDLVQKRLNVCBGGSDGDLFLPPMIRIKGYKVDIESEFFEDVLYNCKPMAS
 GIGIEISTFPEVCTIHASRDRCGCVGTYLIERKQIEGSLDQITLLRCMSVGH
 KYAGLLISCTPSAGKITGMNMYGVTPDDVDELLDQIAGEIITQITPMRQCMIPGEE
 AEKHEQKIVNGHGVKVEESKPTGGCCGCSNVSQCDQETPRPEPIKKEVKCTIIM
 FQPLDKELTYIGAIVGCAIATINAAFTPEKSG"
 complement(join(6276...6315,6416...6509,6717...8472,

CDS

CDS

```
10076..10170,10260..10366,10460..10688,10790..10861)
/note="contains similarity to unknown protein
dbj|BA233705.1
gene_id=MP13_4"
/codon_start=1
/evidence=not_experimental
/protein_id="BAB08523.1"
```

[illegible]

```

23922..24026,24126..24195,24277..24373,24499..24640,
23069..25212)
/note="gb|AAC32067.1
gene_id:MNF13.10"
/codon_start=1
/evidence=not experimental
/product="20S proteasome beta subunit PBB2"
/protein_id="BAB08528.1"
/db_xref="GI:9758084"
/translacion="MSQSYVDIPKGGFSPDLCKRNDMLTOKGLAPSLKTGTTIVG
LIFEDGVILADPATEGPVIADNCKIHYMAPIVSCAGTAADPEAYIDMVSOLD
RLHFGYGRDSRVATLTLLKLHLFSYGWGHAVIAVGWDITGPHLHTLYPHGSTPT
LPFRMTSGSIAMSVFEAKTKBSLTDEGKIYAELACISIFPDLSGSVDVCYT
KGHEYLINTMEPRVTYSKGTSTFKTEVLTLTPILLERIVEAGEAME"
complement(26370..27074)

CDS
alignment_scores:
    Quality: 207.50      Length: 148
    Ratio:   2.530       Gaps:     7
Percent Similarity: 55.405 Percent Identity: 40.541

alignment_block:
US-09-554-547-16 x AB009052 ..
Align seg 1/1 to: AB009052 from: 1 to: 85992

23 ArgLeuGlySerAspLeuAspThrPheGlyArgValHisGlyAsnLeuTy 39
|||||||::: :::::::::::::::::::: ||
1570 CGGTGCCTGACTCCCTCACTTCCTTGCGAGTTAAGAATGCTTA 1619
39 rAlaGlyIleGluArgAlaGlyProArgGlyTyProGlyLeuThrAla 56
|||||::: :::: ||| |||
1620 TGGTGCAGCACACGGTGAAGATGCTGCACAGTGCCTTACAAGT 1669
56 erIleGlyGlyIleValGlyAlaArgLeuGlyArgAlaGlyValGly 72
|||||||::: :::::::::::::::::::: ||
1670 CAGGTGCCGTTATGAGAGGTGGCCGCTCTGCTGTTATGAGATGCGA 1719
73 ValSerSerTyrcGlyTyrcGlyTyrcProserTPGlyTyrcTyrcGly 89
::: |||||::: :::::::::: |||||
1720 GGAGAAATTGGTGGCTATGCCGCTGCTGATATGCT.....GCTGG 1760
89 yTyrcGlyTyrc..... 93
1761 TGTGGACCTTATAGAGAGAACACAGCTCTTGATATTCGTCGTTATG 1810
94 ..Glycyltyrcglytyr....GlycyltyraspIndlyphegly 107
|||||||::: :::::::::::::::::::: ||
1811 GAGGAGAGCGCGGTGTTACAATAAGAGTGATACGATVGGAGAGGCGG 1860
108 SerAlatyrcGlyTyrcProGly...TyrcTyrcGlyTyrcTyrc.... 121
::: |||||::: ||||| |||
1861 GGTGGCTACGCTGCTGACACCAGCTGATATGCTGGGCTCATACGGTCA 1910
122 .....ProserGly...TyrcGlyGlyGlyTyrcG 130
|||||||::: ||||| |||||
1911 ACCAGAGCTGCTTATGCTGACCTAAGTGCAGTTATGAGCGCGGTATG 1960
130 ]y.....GlyserTyrcGlyGlyserTyrcGlyser 140
|| |::: |||||::: |||||:::
1961 GAGTAGACGGAATAGGTGGTATGCTGAGTAATGAGTGGTGGTCT 2004

seq_name: gb_hlg:AC017979

seq_documentation_block:
LOCUS AC017979 23547 bp DNA HMG 09-DEC-1999
DEFINITION Drosophila melanogaster, *** SEQUENCING IN PROGRESS ***, In ordered
pieces.
ACCESSION AC017979
VERSION AC017979.1 GI:6553211
KEYWORDS HMG; HTGS_PHASE2.
SOURCE fruit fly.
ORGANISM Drosophila melanogaster

```

```

REFERENCE: Eurytota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
AUTHORS: Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
TITLE: Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
JOURNAL: 1 (bases 1 to 23547)
          Adams, M. and Venter, J. C.
          Direct Submission
          Submitted (09-DEC-1999) Celera Genomics, 45 West Gude Drive,
          Rockville, MD, USA
COMMENT: This sequence was identified as CDM:10212867 by the submitter.
          For more information on this record e-mail to fly@celera.com.
          * NOTE: This is a 'working draft' sequence.
          * This sequence will be replaced
          * by the finished sequence as soon as it is available and
          * the accession number will be preserved.
FEATURES
  source
    1..23547
    /organism="Drosophila melanogaster"
    /db_xref="taxon:7227"
BASE COUNT  6844 a 4896 c 4767 g 7040 t
ORIGIN
Alignment_scores:
  Quality: 199.50      Length: 92
  Ratio: 3.440         Gaps: 6
  Percent Similarity: 63.043   Percent Identity: 50.000
alignment_block:
  US-09-554-547-16 x AC017979/rev ..
$ Align seg 1/1 to reverse of: AC017979 from: 1 to: 23547
69 AAlGlyVal...GlyValSerSerTYrGlyTYrGlyTYrProSerTrpG1 84
   |||||      :::::  :::::|||||||:::
18369 GCGGAGTGTGTGATGCGCCCGCCACTTGTGATGATGATTCGCA..... 18328
      84 yTYrProTYrGlyGlyTYrGlyTYrGlyGlyTYrGlyGly..... 98
      |||||      |||||      |||||      |||||      |||||
18327 ...CCATATGGAGATACGAGAGATACGAGAGATATGAGGAAGCTATG 18282
      99 .....
      TyGlyGlyTYrAspGinglyPheGly 107
      |||||      |||||      |||||
18281 CCHCGCGCTCTGCCAGCAGCAGTGGCGGAGAGATAT...CAAGATTCGGA 18235
      108 SerAlaTYrGlyGlyTYrProGlyTYrTYrGlyTYrTYrProSerG1 124
      ||  |||||  |||||  |||||  |||||
18234 .....TACCAGAGATACGGGGGATATGAGAGATAC.....GGAGC 18200
      124 yTYrGlyGlyGlyTYrGlyGlySerTYrGlyGlySerTYrGlyGlySert 141
      ||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
18199 AATTGTGTGGAGCATACGGTGGAGCTATGCGGACACATATACACAGCTTTA 18150
      141 yrThTYrProAnValArgAlaSer 149
      ::|||  |||  ::|||
18149 GCACACTACACACTACGATCTTCT 18124
seq_name: gb_in:AC009461
seq_documentation_block:
LOCUS      AC009461      171612 bp      DNA      INV      23-MAR-2001
DEFINITION Drosophila melanogaster, chromosome 3R, region 90B-90C, BAC clone
            BACR1113, complete sequence.
ACCESSION  AC009461
VERSION    AC009461.5  GI:13435227
KEYWORDS   HTG.
SOURCE     fruit fly.
ORGANISM   Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
            1 (bases 1 to 171612)
            Celisner, S.E., Adams, M.D., Kronmiller, B., Tyler, D., Wan, K.H.,
            Holt, R.A., Evans, C.A., Goeayne, J.D., Amandatides, P.G., Brandon, R.C.,

```

TITLE
JOURNAL
REFERENCE
AUTHORS

Rogers, Y., An, H., Baldwin, D., Bazzone, J., Beeson, K.Y., Busan, D.A., Carlson, J.M., Center, A., Champagne, M., Davenport, L.B., Dietz, S.M., Dodson, K.K., Dorsett, V., Doup, L.E., Doyle, C., Dresnek, D., Farfan, D., Ferreira, S., Frick, E., Galle, R.F., Gary, N.S., George, R.A., Gonzalez, M., Houze, J., Hoskins, R.A., Hostin, D., Howland, T.J., Ibdagwam, C., Jalali, M., Murphy, D., Li, P., Mattei, B., Moshrefi, A., McIntosh, T.C., Moy, M., Murphy, B., Nelson, C., Nelson, K.A., Nunoo, J., Pacleb, J., Paradis, V., Park, S., Patel, S., Pfeiffer, B., Phouamavong, S., Plitman, G.S., Pull, V., Richards, S., Scheeler, F., Stapleton, M., Strong, R., Svrtskas, R., Teclor, C., Williams, S.M., Zaveri, J.S., Smith, H.O., Rubin, G.M. and Venter, J.C.

Sequencing of Drosophila chromosome 3R, region 90B-90C

Unpublished

2 (bases 1 to 171612)

Celniker, S.E., Agbayani, A., Arcaina, T.T., Baxter, E., Blazek, R.G., Butenhoff, C., Champe, M., Chavez, C., Chew, M., Ciesiolka, L., Doyle, C.M., Farfan, D.E., Galle, R., George, R.A., Harris, N.L., Hoskins, R.A., Houston, K.A., Humstall, S.R., Karra, K., Kearney, L., Kim, E., Lee, B., Lewis, S., Li, P., Lomoton, M.A., Mazda, P., Moshrefi, A.R., Moshrefi, M., Nixon, K., Pacleb, J.M., Park, S., Pfeiffer, B., Poon, L., Sequella, A., Sethi, H., Snit, E., Svrtskas, R.R., Wan, K.H., Weinburg, T., Zhang, R., Zieran, L.L. and Rubin, G.M.

TITLE
JOURNAL
COMMENT

Submitted (23-AUG-1999) Drosophila genome center, Lawrence Berkeley Laboratory, MS 64-121, Berkeley, CA 94720, USA

On Mar 23, 2001 this sequence version replaced gi:5836088.

Sequence submitted by:
Berkeley Drosophila Genome Project
Lawrence Berkeley National Laboratory, MS 64-121
Berkeley, CA 94720

This sequence was assembled using end sequences from a whole genome shotgun and from subclones of this BAC and its neighboring clones. For further information about this sequence, including its location and relationship to other sequences, please visit our sequence archive Web site (<http://www.fruitfly.org/sequence/>) or send email to bdb@fruitfly.berkeley.edu.

FEATURES
Source

1..171612
/organism="Drosophila melanogaster"
/strain="y; cn bw sp"
/db_xref="taxon:7227"
/chromosome="3R"
/map="90B-90C"
/clone="BACR1L13 (0981)"
/clone_id="RPCI-98 (Rosewell Park Cancer Institute
Drosophila melanogaster BAC library, partial ECORI in
PBacc3.6)"

BASE COUNT 50924 a 36009 c 35092 g 49587 t

ORIGIN

alignment_scores:

	Quality:	Length:
Ratio:	199.50	92
Percent Similarity:	3.440	Gaps: 6
	63.043	Percent Identity: 50.000

alignment_block:
US-09-554-547-16 x AC009461 ..

Align seg 1/1 to: AC009461 from: 1 to: 171612

69 AlaGlyVal...glyAlaSerSerTyrglyTyrglyTyrProserTrpgl 84
||||||| :::: :::::::::::::::
112506 GCCGGAGTGTTAGTCCGCCCGCTTGGATTGATGCACA..... 112547

84 yTyrrProTyrglyGlyTyrglyGlytyrGlyGlyTyrglyGly..... 98
||||||| :::: :::::::::::::::
112548CCATATGAGGATACGAGATACGAGATATGTGGAGAAGCTAAG 112593

99TyrglyGlyTyrraspGnglyPhegily 107
||||||| :::: :::::::::::::::
112594 CCTCCGCTTCGACCAGCACTGCGGAGAGATAT...CAAGATTTCGGA 112640

108 SerAlaTyrGlyGlyTyrProGlyTyrGlyTyrProSerG1 124
112641TACCAGGATACGGGCGATATGAGGATAC.....CGAGG 112675
124 YrTyrGlyGlyGlyTyrGlyGlySerTyrGlyGlySerTyrGlyGlySert 141
112676 ATTTGGTGGAGATACGATACGCTATGGGCGACCAATACACCAAGTTTAA 112725
141 YrTyrTyrProAsnValAlaGAlaSer 149
112726 GCACACTACACCACTACGATCTTCT 112751
seq_name: gb_in:AC009742
seq_documentation_block:
LOCUS AC009742 178100 bp DNA INV 14-MAR-2001
DEFINITION Drosophila melanogaster, chromosome 3R, region 90C-90C, BAC clone
ACCESSION BACR21K03, complete sequence.
VERSION AC009742
KEYWORDS AC009742.4 GI:13324758
SOURCE HTG.
ORGANISM fruit fly.
Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 178100)
Celiker,S.E., Adams,M.D., Krommiller,B., Tyler,D., Wan,K.H.,
Holt,R.A., Evans,C.A., Cocayne,J.D., Amanatides,P.G., Brandon,R.C.,
Rogers,Y., An,H., Baldwin,D., Banzon,J., Beeson,K.Y., Busam,D.A.,
Carlson,J.W., Center,A., Champagne,M., Davenport,L.B., Dietz,S.M.,
Dodson,K., Dorsett,V., Doup,L.E., Doyle,C., Dreene,D., Farfan,D.,
Ferreira,S., Frise,E., Galle,R.F., Garg,N.S., George,R.A.,
Gonzalez,M., Houck,J., Hoskins,R.A., Hostin,D., Howland,T.J.,
Ibegwam,C., Jalali,M., Kruse,D., Li,P., Matczi,B., Moshrefi,A.,
McIntosh,T.C., Moy,M., Murphy,B., Nelson,C., Nelson,K.A., Nunoo,J.,
Pachleb,J., Paragas,V., Park,S., Patel,V., Pfeiffer,B.,
Phouanavong,S., Piltman,G.S., Puri,V., Richards,S., Scheeler,F.,
Stapleton,M., Strong,R., Svrtskas,R., Tector,C., Williams,S.M.,
Zaveri,J.S., Smith,H.O., Rubin,G.M. and Venter,J.C.
Sequencing of Drosophila chromosome 3R, region 90C-90C
Unpublished
2 (bases 1 to 178100)
TITLE
JOURNAL
REFERENCE
AUTHORS
Celiker,S.E., Abdayani,A., Arcalata,T.T., Baxter,E., Blazej,R.G.,
Butenhoff,C., Champagne,M., Chavez,C., Chew,M., Ciesiolka,L.,
Doyle,C.M., Farfan,D.E., Galle,R., George,R.A., Harris,N.L.,
Hoskins,R.A., Houston,K.A., Hummasti,S.R., Karra,K., Kearney,L.,
Kim,E., Lee,B., Lewis,S., Li,P., Lomolan,M.A., Mazda,P.,
Moshrefi,A.R., Moshrefi,M., Nixon,K., Pachleb,J.M., Park,S.,
Pfeiffer,B., Poon,L., Sequelira,A., Sethi,H., Shiri,E.,
Svrtskas,R.R., Wan,K.H., Weinburg,T., Zhang,R., Zieran,L.L. and
Rubin,G.M.
Direct Submission
Submitted (30-AUG-1999) Drosophila Genome Center, Lawrence Berkeley
Laboratory, MS 64-121, Berkeley, CA 94720, USA
COMMENT
Sequence submitted by:
On Mar 14, 2001 this sequence version replaced gi:5922029.
Berkeley Drosophila Genome Project
Lawrence Berkeley National Laboratory, MS 64-121
Berkeley, CA 94720
This sequence was assembled using end sequences from a whole genome
shotgun and from subclones of this BAC and its neighboring clones.
For further information about this sequence, including its location
and relationship to other sequences, please visit our sequence
archive Web site (<http://www.fruitfly.org/sequence/>) or send email
to bdgfruitfly@berkeley.edu.
FEATURES
source
Location/Qualifiers
1. 178100
/organism="Drosophila melanogaster"
/strain="Y: cn bw sp"
/db_xref="taxon:7227"
/chromosome="3R"

/map="90C-90C"
/clone="BACR21K03 (D982)"
/clone_lib="RPCI-98 (Foswell Park Cancer Institute
Drosophila melanogaster BAC library, partial EcoRI in
pBAC3.6)"
BASE COUNT 51308 a 37603 c 37904 g 51285 t
ORIGIN
alignment_scores:
Quality: 199.50 Length: 92
Ratio: 3.440 Gaps: 6
Percent Similarity: 63.043 Percent Identity: 50.000
alignment_block:
US-09-554-547-16 x AC009742 ..
Align seq 1/1 to: AC009742 from: 1 to: 178100
69 AlaglyVal...GlyValSerSerTyrGlyTyrGlyTyrProSerTyrG1 84
47019 GCGGAGGTGTGAGTGTGCGCCCGAGTTGATATGATTCGCA..... 47060
84 YrTyrProTyrGlyGlyTyrGlyGlyTyrGlyGlyTyrGlyGly..... 98
47061CCATATGAGAGATACGAGAGATACGAGATATGAGGAAGCTATG 47106
99TyrGlyGlyTyrAspGlnGlyPheGly 107
47107 CTTCCGCCCTTCCGACGACGATGCGGAGGATAT...CAAGATTTGGA 47153
108 SerAlaTyrGlyGlyTyrProGlyTyrGlyTyrTyrTyrProSerG1 124
47154TACCAGGATACGGGCGATATGAGGATAC.....GGAGG 47188
124 YrTyrGlyGlyGlyTyrGlyGlySerTyrGlyGlySertTyrGlySert 141
47189 ATTTGGTGGAGATACGCTATGGGCGACCAATACACCAAGTTTAA 47238
141 YrTyrTyrProAsnValAlaGAlaSer 149
47239 GCACACTACACCACTACGATCTTCT 47264
seq_name: gb_in:AE003718
seq_documentation_block:
LOCUS AE003718 253176 bp DNA INV 05-OCT-2000
DEFINITION Drosophila melanogaster genomic scaffold 142000013386035 section 43
OF 105, complete sequence.
ACCESSION AE003718 AE002708
VERSION AE003718.2 GI:10726582
KEYWORDS HTG.
SOURCE fruit fly.
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 253176)
Adams,M.D., Celiker,S.E., Holt,R.A., Evans,C.A., Cocayne,J.D.,
Amanatides,P.G., Scherer,S.E., Li,P.W., Hoskins,R.A., Galle,R.F.,
George,R.A., Lewis,S.E., Richards,S., Ashburner,M., Henderson,S.N.,
Sutton,G.G., Wortman,J.R., Yandell,M.D., Zhang,Q., Chen,L.X.,
Brandon,R.C., Rogers,Y.H., Blazej,R.G., Champagne,M., Pfeiffer,B.D.,
Wan,K.H., Doyle,C., Baxter,E.G., Helt,G., Nelson,C.R., Gabor
Mikos,G.L., Abriil,J.F., Abdayani,A., An,H.J.,
Andrews-Frankkoci,C., Baldwin,D., Ballew,R.M., Basu,A.,
Baxendale,J., Bayraktaroglu,L., Beasley,E.M., Beeson,K.Y.,
Benos,P.V., Berman,B.P., Bhandari,D., Bolshakov,S., Borkova,D.,
Botchan,M.R., Bouck,J., Brokstein,P., Brothier,P., Butlis,K.C.,
Busam,D.A., Butler,H., Cadieu,E., Center,A., Chandra,I.,
Cherry,J.M., Cawley,S., Dahlke,C., Davenport,L.B., Davies,P., de
Pablos,B., Delcher,A., Deng,Z., Mays,A.D., Dew,I., Dietz,S.M.,
Dodson,K., Doup,L.E., Downes,M., Dugan-Rocha,S., Dunkov,B.C.,

Dunn, P., Durbin, K.J., Evangelista, C.C., Ferraz, C., Ferreira, S.,
Fleischmann, W., Foster, C., Gabriellian, A.E., Garg, N.S.,
Gelbart, W.M., Glasser, K., Glodzik, A., Gong, F., Gorrell, J.H., Gu, Z.,
Guan, P., Harris, M., Harris, N.L., Harvey, D., Helman, T.J., Hu, Z.,
Hernandez, J.R., Houck, J.A., Hostin, D., Houston, K.A., Howland, T.J.,
Jell, M.H., Idegiam, C., Jaitani, M., Kalush, F., Karp, G.H., Ke, Z.,
Kennell, J.A., Ketchum, K.A., Kimmel, B.E., Kodira, C.D., Kraft, C.,
Krawitz, S., Kuip, D., Lai, Z., Laekko, P., Lei, Y., Levitsky, A.A.,
Li, J., Li, Z., Liang, Y., Lin, X., Liu, X., Maltel, B., McIntosh, T.C.,
McLeod, M.P., McPherson, D., Merkulov, G., Mishina, N.V., Mobarry, C.,
Morris, J., Moshrefi, A., Mount, S.M., Moy, M., Murphy, B., Murphy, L.,
Muzny, D.M., Nelson, D.L., Nelson, D.R., Nelson, K.A., Nixon, K.,
Nuzny, D.M., Pacle, J.W., Palazzolo, M., Peltman, G.S., Pan, S.,
Pollard, J.D., Puri, V., Reese, M.G., Reinert, K., Remington, K.,
Sander, R.D., Scheeler, F., Shen, H., Shue, B.C., Siden-Kiamos, I.,
Simpson, M., Skupski, M.P., Smith, T., Spier, E., Spector, C., Turner, R.,
Stapleton, M., Strong, R., Sun, E., Switzkas, R., Tector, C., Turner, R.,
Venter, E., Wang, A.H., Wang, X., Wang, Z., Wasserman, D.A.,
Weinstock, G.M., Weissbach, J., Williams, S.M., Woodage, T.,
Worley, K.C., Wu, D., Yang, S., Yao, Q.A., Ye, J., Yeh, R.F.,
Zaveri, J.S., Zhao, M., Zhang, G., Zhao, Q., Zheng, L., Zheng, X.H.,
Zhong, F.N., Zhong, W., Zhou, X., Zhu, S., Zhu, X., Smith, H.O.,
Gibbs, R.A., Myers, E.W., Rubin, G.M. and Venter, J.C.
The genome sequence of *Drosophila melanogaster*
Science 287 (5461), 2185-2195 (2000)
2 (bases 1 to 253176)
20196006
JOURNAL
MEDLINE
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
FEATURES
source
On Oct 9, 2000 this sequence version replaced gi:7300262.
Location/Qualifiers
1. 253176
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/chromosome="3R"
complement(join(<9015..9066,9230..9493,9803..10067,
10198..10425,10609..10857,11045..11208,11695..12010,
12723..12848,13260..13634,13863..14042,14361..14516,
14593..14901,15656..15799,15889..16144,16308..16518,
16861..17087,17202..17380,17568..17597,17972..18091,
18411..18546,18901..19046,19141..19290,19363..19494,
21408..21612,21940..22040,22140..22267,23854..24077,
24436..24553,25877..25997,26180..26329,26586..26747,
27102..>27144))
/gene="CG12536"
/product="CT33960"
/db_xref="FLYBASE:FBan0012536"
/db_xref="FLYBASE:FBgn0038513"
complement(<9015..>27144)
/gene="CG12536"
/map="90B1-90B2"
/db_xref="FLYBASE:FBan0012536"
/db_xref="FLYBASE:FBgn0038513"
complement(join(9015..9066,9230..9493,9803..10067,
10198..10425,10609..10857,11045..11208,11695..12010,
12723..12848,13260..13634,13863..14042,14361..14516,
14593..14901,15656..15799,15889..16144,16308..16518,
16861..17087,17202..17380,17568..17597,17972..18091,
18411..18546,18901..19046,19141..19290,19363..19494,
21408..21612,21940..22040,22140..22267,23854..24077,
24436..24553,25877..25997,26180..26329,26586..26747,
27102..27144))
/gene="CG12536"
/note="CG12536 gene product"
/codon_start=1
/db_xref="FLYBASE:FBan0012536"
/db_xref="FLYBASE:FBgn0038513"
/protein_id="AAFS5425.1"
/db_xref="GI:7300263"
/translation="MRLRRAGSSTAVLRQFHVHIVENTEVYLSALIKAIPEY
VRBYVAVSWHRGEETLLPDLSDVAGRYVLAASGLYVRSVSESDGLMKFSLVNT
LNGRSDAVALQVKLSKNLAPRTQKQVMEIHERGNDVILPCIQNPPIETW
YRVSDAALPIPSQVILSRITLLIKMDEPRAGWICQASQFEEDRIERLSYN
SYVSHILPOVQISNGTANFNCTLLSIDAIDWLNKRPLODANNAITGRNIF
LSKSLSLVAVNGRQVOCYLVNORASQAMKELIGTPEVLTFTLEONRPP
LISKCSAGSPRRQFAMLDSDOPIVMISLHHPRAICQFQMSGDVLSHINVRPD
DGCIYKVASNSKQSVSHARLANYGPYRALCPKAVAGEDITVCPRAGYVEOI
RNEAHQELTTSNHYELASVADGGLVIRNPERQDGIITCYRSGABERDOL
KWSAPVIEPFEPKRNLOEGRAITAVSSGDPIYFSMKKDDSSIPSSQITERKE
EYSLVLEKDISARHSKRYTCYSNAAKVNYAVLOVRAPRRYRPPMTALMGIT
ISINCEAGYPIPIITWFGKSGKSKDFKLPARFQOSARINISRRNDVTLCHK
GDEPITIGMONGRIDLNRFSLIAKTEKGERPTPSLEIFEEFSRYKLSMR
PFGNSVLYLVQYQALYKIQSGSLAAGCMNGVINYISPTSEISYSDLE
SAIVAGLTPATTEILRMQALNELERSATELVKTOEAPTEAPSVVOYQGESSEL
IYVQIPIPRSMWELIGYVNSCEEQUNIFISVNSSLKSTIVSGWATKATRL
KYSVAPVITIRAMSGFGSPSAIFGTEGEVAPQVNCALTSOSIKTSLMR
PLOGHIGIQCYKILRYPIVQIDPEPKLTKRPSNLEYLHTLTKRASNISRYLAT
ATGGLASHPLFCOTDDVDVPAAPAKAALATASIIISMLTPKRNGLTISHYVYR
EAGKCGAKTHMAYVDNGPYTESLSLENQYERWVSASTVSGCEPRTVYAQT
NTRAPARIASFQVVRKAVGGLVLECLAVGNPPIRKMVLTRPVPYTESVNBG
NLKIHREGSLSGVYCTANLFGSDIENQVIAKPPSAQIIVQASADISVND
APDGCAPLQGYTISYTAGESWITELLENNAFTISGLCGQYIIKMSHNVGS
GVASEELNWTMRKASQAPANANELIATNATCNVLSKLSMONGGCSIHFESEHPLD
IRMTYVSDISNAEENELIFCDLPAKAWQLASIPNAGKTEHYHSTTIDGI
TIIPPVSFPESENDMMNLINSTNPTSGDWATLIVYITYSITITLTKHRTLCG
PIAGYESRPLPGDYKEDHENRQOVSAPVTVKGNSEMEITSPATKESVNG
RTGAPAKTPRAVAQPTPLDTMQKFTGHEGELNATAVAPLPSGFGVKKSSM
HKORYVNTEDESTLSKMTIYAGSQASHKSNKSGRASKSAGVAGGSESTIS
PSTEFNMPTRYRVCSSRSRSDGRAVADQRPDSSTSNNOGSPAPERIRNTRHVL
CMGAMGLGGGCGGCGNGCPARKRAGQSRKRGSAQOQSNQTLERRKPCSSNS
LDSEVDNETASLAASTAAAGSAGADLSGCFRPPAGFHGGRPGHSDCERQAL
DLEQVWSESTGDMQLMKREELTSLARHEKKEDERLTTHV"
complement(join(31465..31726,32418..32457,
39745..39767))
/gene="CG8106"
/product="CT24226"
/db_xref="FLYBASE:FBan0008106"
/db_xref="FLYBASE:FBgn0038514"
complement(<31465..>39767)
/gene="CG8106"
/map="90B1-90B3"
/db_xref="FLYBASE:FBan0008106"
/db_xref="FLYBASE:FBgn0038514"
complement(join(31503..31726,32418..32457))
/gene="CG8106"
/note="CG8106 gene product"
/codon_start=1
/db_xref="FLYBASE:FBan0008106"
/db_xref="FLYBASE:FBgn0038514"
/protein_id="AAFS5426.1"
/db_xref="GI:7300264"
/translation="MSLLEPCLVLEKALLGALLAVIATPLATSGLRVPTLLEP
APRLFGNDCAQVTCYAHGSPPLVWVLPDGLAVQVPLR"
complement(join(66536..67410,67471..67541,67834..68268,
68330..68753))
/gene="CG5823"
/product="CT18279"
/db_xref="FLYBASE:FBan0005823"
/db_xref="FLYBASE:FBgn0038515"
complement(<66536..>68753)
/gene="CG5823"
/map="90B3-90B4"
/db_xref="FLYBASE:FBan0005823"
/db_xref="FLYBASE:FBgn0038515"
complement(join(66931..67410,67471..67541,67834..68134))
/gene="CG5823"
/note="CG5823 gene product"
/codon_start=1
/db_xref="FLYBASE:FBan0005823"
/db_xref="FLYBASE:FBgn0038515"
/protein_id="AAFS5427.2"
/db_xref="GI:10726583"
/translation="MSSSYHSGGRKQPTAVSRMKQDYWRKLRDPLPYITAPLPDNL
EMHYCVKPEDSPYGYHTLLPREFPEPKPPSIYMLPNGRFRKTNTRICLSISDF

LNGRSDAVALQVKLSKNLAPRTQKQVMEIHERGNDVILPCIQNPPIETW
YRVSDAALPIPSQVILSRITLLIKMDEPRAGWICQASQFEEDRIERLSYN
SYVSHILPOVQISNGTANFNCTLLSIDAIDWLNKRPLODANNAITGRNIF
LSKSLSLVAVNGRQVOCYLVNORASQAMKELIGTPEVLTFTLEONRPP
LISKCSAGSPRRQFAMLDSDOPIVMISLHHPRAICQFQMSGDVLSHINVRPD
DGCIYKVASNSKQSVSHARLANYGPYRALCPKAVAGEDITVCPRAGYVEOI
RNEAHQELTTSNHYELASVADGGLVIRNPERQDGIITCYRSGABERDOL
KWSAPVIEPFEPKRNLOEGRAITAVSSGDPIYFSMKKDDSSIPSSQITERKE
EYSLVLEKDISARHSKRYTCYSNAAKVNYAVLOVRAPRRYRPPMTALMGIT
ISINCEAGYPIPIITWFGKSGKSKDFKLPARFQOSARINISRRNDVTLCHK
GDEPITIGMONGRIDLNRFSLIAKTEKGERPTPSLEIFEEFSRYKLSMR
PFGNSVLYLVQYQALYKIQSGSLAAGCMNGVINYISPTSEISYSDLE
SAIVAGLTPATTEILRMQALNELERSATELVKTOEAPTEAPSVVOYQGESSEL
IYVQIPIPRSMWELIGYVNSCEEQUNIFISVNSSLKSTIVSGWATKATRL
KYSVAPVITIRAMSGFGSPSAIFGTEGEVAPQVNCALTSOSIKTSLMR
PLOGHIGIQCYKILRYPIVQIDPEPKLTKRPSNLEYLHTLTKRASNISRYLAT
ATGGLASHPLFCOTDDVDVPAAPAKAALATASIIISMLTPKRNGLTISHYVYR
EAGKCGAKTHMAYVDNGPYTESLSLENQYERWVSASTVSGCEPRTVYAQT
NTRAPARIASFQVVRKAVGGLVLECLAVGNPPIRKMVLTRPVPYTESVNBG
NLKIHREGSLSGVYCTANLFGSDIENQVIAKPPSAQIIVQASADISVND
APDGCAPLQGYTISYTAGESWITELLENNAFTISGLCGQYIIKMSHNVGS
GVASEELNWTMRKASQAPANANELIATNATCNVLSKLSMONGGCSIHFESEHPLD
IRMTYVSDISNAEENELIFCDLPAKAWQLASIPNAGKTEHYHSTTIDGI
TIIPPVSFPESENDMMNLINSTNPTSGDWATLIVYITYSITITLTKHRTLCG
PIAGYESRPLPGDYKEDHENRQOVSAPVTVKGNSEMEITSPATKESVNG
RTGAPAKTPRAVAQPTPLDTMQKFTGHEGELNATAVAPLPSGFGVKKSSM
HKORYVNTEDESTLSKMTIYAGSQASHKSNKSGRASKSAGVAGGSESTIS
PSTEFNMPTRYRVCSSRSRSDGRAVADQRPDSSTSNNOGSPAPERIRNTRHVL
CMGAMGLGGGCGGCGNGCPARKRAGQSRKRGSAQOQSNQTLERRKPCSSNS
LDSEVDNETASLAASTAAAGSAGADLSGCFRPPAGFHGGRPGHSDCERQAL
DLEQVWSESTGDMQLMKREELTSLARHEKKEDERLTTHV"
complement(join(31465..31726,32418..32457,
39745..39767))
/gene="CG8106"
/product="CT24226"
/db_xref="FLYBASE:FBan0008106"
/db_xref="FLYBASE:FBgn0038514"
complement(<31465..>39767)
/gene="CG8106"
/map="90B1-90B3"
/db_xref="FLYBASE:FBan0008106"
/db_xref="FLYBASE:FBgn0038514"
complement(join(31503..31726,32418..32457))
/gene="CG8106"
/note="CG8106 gene product"
/codon_start=1
/db_xref="FLYBASE:FBan0008106"
/db_xref="FLYBASE:FBgn0038514"
/protein_id="AAFS5426.1"
/db_xref="GI:7300264"
/translation="MSLLEPCLVLEKALLGALLAVIATPLATSGLRVPTLLEP
APRLFGNDCAQVTCYAHGSPPLVWVLPDGLAVQVPLR"
complement(join(66536..67410,67471..67541,67834..68268,
68330..68753))
/gene="CG5823"
/product="CT18279"
/db_xref="FLYBASE:FBan0005823"
/db_xref="FLYBASE:FBgn0038515"
complement(<66536..>68753)
/gene="CG5823"
/map="90B3-90B4"
/db_xref="FLYBASE:FBan0005823"
/db_xref="FLYBASE:FBgn0038515"
complement(join(66931..67410,67471..67541,67834..68134))
/gene="CG5823"
/note="CG5823 gene product"
/codon_start=1
/db_xref="FLYBASE:FBan0005823"
/db_xref="FLYBASE:FBgn0038515"
/protein_id="AAFS5427.2"
/db_xref="GI:10726583"
/translation="MSSSYHSGGRKQPTAVSRMKQDYWRKLRDPLPYITAPLPDNL
EMHYCVKPEDSPYGYHTLLPREFPEPKPPSIYMLPNGRFRKTNTRICLSISDF

HPDVMNPFVWCVGTILTGLLSPMLESTPPLGSTEESNYDKQFAKSLAFNIRNPNFEE
LEPELVEIEIKORLRTGTOAAAADVGPAAAGVRSKGLAGASLASADGLNVLGGLANDE
NAADAIDGGAAGQADLHAGGCAAVQNSARNYILMNQSYNSLVITICEFALFALIVNV
IKNLQGE*

Join(69939, 70062, 70244, 70909, 70982, ..>71058)

mirna


```

|||||
97  TGGAGCAGAGGCTGCTGCTTGGAGACGACCTGGATACATTTGGAGGG 146
34  AHHHSGIYASnLeuTyraIaGlyIleGluArGaIaGlyProArGlyTyrr 50
|||||
147 TACACGGTATACCTATATGCGGCATCGAAGAGCTGCGCTCGTGATAC 196
51  ProGlyLeuThraIaSerIleGlyIaGlyIaGlyIaArGleuGlyI 67
|||||
197 CAGAGGCTTACCGCATTCGATTGGAGGGAGTGGTGCACAGACTCGGTGG 246
67  YArGaIaGlyIaGlyIaSerSerTyrrGlyTyrrProSerTrpG 84
|||||
247 TCGTGCCGGTGTGGAGTACGACCTACGGCTATGTTACCTTCATGGG 296
84  LyrrProTyrrGlyIaGlyTyrrGlyIaGlyTyrrGlyTyrrGly 100
|||||
297 GCTATCCGATAGGTGATACGGTGGATACGGTGGATACGGT 346
101 GLyTyrrAspInGlyPheGlySerAlaTyrrGlyIaGlyTyrrProGlyTyrr 117
|||||
347 GGAATATACAGGGTTTGGCTCTGCATACGGCGGCTACCGCGCTACTA 396
117 rGlyTyrrTyrrProSerGlyTyrrGlyIaGlyTyrrGlyIaSerTyrrG 134
|||||
397 TGGCTACTACTATCCAGTGGCTACGGTGGGCTACGGTGGTACG 446
134 LyGlySerTyrrGlyIaSerTyrrThrTyrrProAsnValArGaIaSerAla 150
|||||
447 GTGGCACTACGGTGGTACGCTACACTATCCCAACGTTCCGGCTTCAGCT 496
151 GLyAlaIaAla 154
|||||
497 GGTGCCCGCAGCT 508

```

seq_name: /stds/gcgdata/geneseq/geneseqn/NA2000.DAT:AAA64661

seq_documentation_block:
ID AAA64661 standard; cDNA: 2126 BP.

AC AAA64661;

DT 02-JAN-2001 (first entry)

DE cDNA encoding TBP associated factor (TAFII68).

KW Melanoma: thyroid tumour; MEL3; MEL7; Thy5; Thy6; Thy11; Thy14; Thy15;

KW rectal cancer; lung cancer; breast cancer; colon cancer;
TBP associated factor; TAFII68; ss.

OS Homo sapiens.

PN W0200050595-A2.

XX 31-AUG-2000.

XX 25-FEB-2000: 2000WO-US04929.

XX 25-FEB-1999: 99US-0257417.

XX (GOUT/) GOUT I.

XX (RODN/) RODIN N.

XX (FILO/) FILOENKO V.

XX (MATS/) MATSUKA G.

XX (SCAN/) SCANLAN M.

XX (OLDL/) OLD L.

XX (BILY/) BILYNSKY B.

XX Gout I, Rodin N, Filonenko V, Matsuka G, Scanlan M, Old L,
XX Bilynsky B;
XX WPI: 2000-572092/53.
XX DR

PT Novel isolated nucleic acid molecules for diagnosing and treating
PT melanoma, thyroid tumors, rectal, lung, breast and colon cancers
XX
XX
PS Claim 17: Page 88; 94pp; English.

CC The specification describes polynucleotides which are associated with
CC melanoma and thyroid tumors. Specifically, these are MEL3, MEL7, Thy5,
CC Thy6, Thy11, Thy14, and Thy15. The polynucleotides are useful for
CC diagnosing and treating a patient with melanoma, thyroid tumour,
CC rectal cancer, lung cancer, breast cancer or colon cancer. The
CC present sequence represents a polynucleotide of the invention.

SQ Sequence 2126 BP; 609 A; 373 C; 691 G; 453 T; 0 other;

alignment_scores:
Quality: 202.50 Length: 138
Ratio: 2.531 Gaps: 9
Percent Similarity: 57.971 Percent Identity: 43.478

alignment_block:
US-09-554-547-16 x AAA64661 ..

Align seg 1/1 to: AAA64661 from: 1 to: 2126

```

20  ArGaIaGlyArGleuGlySerAspLeuAspThrPheGlyArGaValHISG1 36
   ||::|||  ||::|||
1412 AGAGCGCGCGCTATGCGGGGAC.....AGAGG 1440
36  YAsnLeuTyraIaGlyIleGluArGaIaGlyProArGlyTyrrProGlyL 53
   ::::  ||::|||  ::::|||  |||
1441 AGCGCGCTATGAGGA...GACCGAGAGAGT.....GGCTATGAGAG 1481
53  eutHraIaSerIleGlyIaGlyIaValGlyIaArGleuGlyGly...Arg 68
   ::::  ||::|||  ::::|||  |||
1482 ATCGAGGTGGCTATGAGAGAGACCGAGGTGGAGCTATGCTGAGAGACCGA 1531
69  ALaGlyAlaGlyIaSerSerTyrrGlyTyrrGlyTyrrProSerTyrrGlyTy 85
   ::|||  ||  ::  ||::|||  ::  ||::|||
1532 GAGGCTATGAGAGAGATCGAGAGCTATGAGAGACGCGGGGCTATG 1581
1582 TGGAGGAGATCGAGAGAGCTATGAGAGACGAGAGACGCGGGGCTATG 1631
85  T.....ProTyrrGlyTyrrGlyTyrrGlyTyrrGlyTyrrG 100
   ||::|||  ||  ::  ||::|||  ::  ||::|||
100  LyGlyTyrrAspInGlyPheGlySerAlaTyrrGlyTyrrGlyTyrrG 114
   ||||  ||::|||  ||||  ::|||  ::|||  ::|||
1632 GAGGA...GACCGGTGGTGGCAGTGGTGGACACCGAGAGTGA 1678
115  GLyTyrrTyrrGlyTyrrTyrrProSerGlyTyrr.....GlyG1 127
   ||||  ||  ::  ||::|||  ::  ||::|||
1679 GCGTATGAGAGACGACGACGCTGCGCGCTATGAGAGACGACGAGTGG 1728
127  yGlyTyrrGly.....GlySerTyrrGlyGlySerTyrrGlyGlySerT 141
   ||||  ||||  ||::|||  ||::|||  ::  ||||
1729 GGGCTACGAGAGACCGAGGTGCTATGAGCAAAATGGAGGAGAA 1778
141  yrrHrrTyrrProAsn 145
   ||  ||  ||
1779 ACGACTACAGAAAT 1792

```

seq_name: /stds/gcgdata/geneseq/geneseqn/NA2001.DAT:AA110862

seq_documentation_block:
ID AA110862 standard; DNA: 479 BP.

AC AA110862;

DT 12-OCT-2001 (first entry)

DE Probe #795 for gene expression analysis in human cervical cell sample.

KW Probe: human; microarray; gene expression; cervical epithelial cell;

XX cervical cancer; ss.
OS Homo sapiens.
PN MO200157278-A2.
PD 09-AUG-2001.
XX
XX 30-JAN-2001; 2001WO-US00670.
XX
XX 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000CB-0024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
P1 Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR WPI; 2001-488901/53.
XX
PT Human genome-derived single exon nucleic acid probes useful for
XX analyzing gene expression in human cervical epithelial cells -
PS
XX Claim 25; SEQ ID NO 795; 487bp; English.
CC
CC The present invention relates to human single exon nucleic acid probes
CC (SENPs). The present sequence is one such probe. The SENPs are derived
CC from human HeLa cells. The SENPs can be used to produce a single exon
CC microarray, which can be used for measuring human gene expression in a
CC sample derived from human cervical epithelial cells. By measuring gene
CC expression, the probes are therefore useful in grading and/or staging
CC of diseases of the cervix, notably cervical cancer.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at fcp.wipo.int/pub/published_pcl_sequences.
XX
XQ Sequence 479 BP; 76 A; 38 C; 227 G; 138 T; 0 other;
SQ

Alignment_scores:
Quality: 180.50 Length: 101
Ratio: 3.167 Gaps: 3
Percent Similarity: 56.436 Percent Identity: 45.545

Alignment_block:
US-09-554-547-16 x AAII0862 ..

Align seg 1/1 to: AAII0862 from: 1 to: 479

39 TYRATAGIYIEGLIUAARGAAGIYPRARFGLTYTYPROGILYLeutThra1 55
|||:::||| ::||| |||
217 TATGTGGTGAGCGTGTGGTCGATGAGATTATGTTGGT..... 258
55 aserleeglygylualglalargleugllycylarfgalaaglyalg 72
259 ...GCTGTGGTGATGAGCATTATGTCGTGGTCGATGATGCTGTGGTG 304
72 LVVALSERSEFYRGITYTFRPROSEFTFGILTYPYROFIYGL 88
|| |||||||| ::| ||| |||||
305 GA.....GCTTATGCTGTGCAGCGTGTGCCGCGAGTTATGCT 342
89 GLTYTFCIGLYLTyrIGLYLTyrISerIGLYLTyrIGLYLTyrASPGLNG1 105
|| |||||| |||||||| ||||| |||
343 GCTGTGCTGTGCATGACGATTATGCTGTGTGTGTGTGATGAGAG 392
105 yphcglyserialtyrIGLYLTyrPRocGLTYTyrIGLYLTyrTYrrp 122
|||||::||| |||||
393 TTATGTGTGTGTGTGTGTGT.....G 415

```

seq_name: /SID58/gcdata/geneseq/genesegn/NA2001.DAT:AAI32122
seq_documentation_block:
ID AAI32122 standard; DNA; 479 BP.
XX
AC AAI32122;
XX
DT 17-OCT-2001 (first entry)
XX
DE Probe #808 used to measure gene expression in human placenta sample.
XX
KW Probe; microarray; human; placenta; antenatal diagnosis;
XX genetic disorder; ss.
XX
OS Homo sapiens.
XX
PN WO200157272-A2.
PD
PP 09-AUG-2001.
XX
PF 30-JAN-2001; 2001WO-US00663.
XX
PR 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR WPI: 2001-488897/53.
XX
PT Human genome-derived single exon nucleic acid probes useful for
PT analyzing gene expression in human placenta -
XX
XX Claim 25; SEQ ID No 808; 654bp; English.
XX
XX The present invention relates to single exon nucleic acid probes (SENP).
XX CC The present sequence is one such probe. The probes are useful for
CC producing a microarray for predicting, measuring and displaying gene
CC expression in samples derived from human placenta. The probes are useful
CC for antenatal diagnosis of human genetic disorders.
XX
SQ Sequence 479 BP; 76 A; 38 C; 227 G; 138 T; 0 other;

alignment_scores:
Quality: 180.50 Length: 101
Ratio: 3.167 Gaps: 3
Percent Similarity: 56.436 Percent Identity: 45.545

alignment_block:
US-09-554-547-16 x AAI32122 ..

Align seg 1/1 to: AAI32122 from: 1 to: 479

39 TYRAlGlyllegluarGlaagProArGlyTYrProGlyLeutHraI 55
|||||::||| ::||| ||||| |
217 TATGTGCGTGGAGAGTGCTGCTGCTATGAGGTTATGTCGCT..... 258
55 aserlleglglgylguValGlAalArgrBenglglYlarGlaaglYaIg 72
```



```
122 roserGlyTyrGlyGlyTyrGlyGlySerTyrGlySerTyrGly 138
      ::::::::::::::::::::
435 ATGAGCTATGCTGCTGCTGCTGCTGCTATGCTGCTGCTGCTGCT 484
139 Gly 139
      |||
485 GGT 487

seq_name: /SID58/gcdata/geneseq/geneseqn/NA2001.DAT:AA107740

seq_documentation_block:
ID   AA107740 standard; DNA; 830 BP.
XX
AC   AA107740;
XX
DT   09-OCT-2001 (first entry)
XX
DE   Probe #7731 used to measure gene expression in human breast sample.
XX
KW   Probe; human; breast disease; breast cancer; development disorder; ss;
XX   inflammatory disease; proliferative breast disease; non-carcinoma tumour.
XX
OS   Homo sapiens.
XX
PN   WO200157270-A2.
XX
PD   09-AUG-2001.
XX
PE   29-JAN-2001; 2001WO-US00661.
XX
PR   04-FEB-2000; 2000US-0180312.
XX   26-MAY-2000; 2000US-0207456.
XX   30-JUN-2000; 2000US-0608408.
XX   03-AUG-2000; 2000US-0632366.
XX   21-SEP-2000; 2000US-0234687.
XX   27-SEP-2000; 2000US-0236359.
XX   04-OCT-2000; 2000GB-0024263.
XX
PA   (MOLE-) MOLECULAR DYNAMICS INC.
XX
PT   Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR   WPI; 2001-476286/51.
XX
PT   Novel single exon nucleic acid probe used to measuring gene expression
XX   in a human breast -
XX
PS   Claim 25; SEQ ID NO 7731; 322pp; English.
XX
CC   The present invention relates to novel single exon nucleic acid probes.
XX   The present sequence is one such probe. The probes are useful for
XX   measuring human gene expression in a human breast sample, where the probe
XX   hybridises at high stringency to a nucleic acid expressed in the human
XX   breast. The probes are useful for predicting, diagnosing, grading,
XX   staging, monitoring and prognosing diseases of the human breast.
XX   Particularly those diseases with polygenic aetiology. The diseases
XX   include: breast cancer, disorders of development, inflammatory diseases
XX   of the breast, fibrocystic changes, proliferative breast disease and
XX   non-carcinoma tumours.
XX
CC   Note: The sequence data for this patent did not form part of the printed
XX   specification, but was obtained in electronic format directly from WIPO
XX   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 830 BP; 193 A; 86 C; 336 G; 215 T; 0 other;

alignment_scores:
      Quality: 180.50          Length: 101
      Ratio: 3.167            Caps: 3
Percent Similarity: 56.436    Percent Identity: 45.545

alignment_block:
US-09-554-547-16 x AA107740
```

```
Align seg 1/1 to: AA107740 from: 1 to: 830

39 TyrAlaGlyIleLeuArgAlaGlyProArgGlyTyrProGlyLeuThrAl 55
      ||:::|||||
236 TATGAGTGTGAGAGTGTGTGTGATGAGGTTATGCTGT. .... 277
55 aserIleGlyGlyIleValGlyValAArgLeuGlyGlyArgAlaGlyValG 72
      |||||::: |||
278 ...GGTGGTGTGATGAGAGTTATGCTGTGTGTGATGAGTGTGTG 323
72 lYvalSerSerTyrGlyTyrProSerTyrpGlyTyrProTyrGly 88
      ||
324 GA.....CGTTATGCTGTGAGAGTGTGTGCGGAGGTTATGCT 361
89 GlyTyrGlyGlyTyrGlyGlyTyrGlyGlyTyrGlyTyrAspGlnG 105
      || ||||| |||||
362 GGTGGTGGTGTGATGAGGTTATGCTGTGTGTGTGATGAGG 411
105 yPheGlySerAlaTyrGlyTyrProGlyTyrGlyTyrTyrTyrP 122
      ||:::||||: |||||
412 TTATGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 434
122 roserGlyTyrGlyGlyTyrGlyGlySerTyrGlyGlySerTyrGly 138
      ::::::::::::::::::::
435 ATGAGCTATGCTGCTGCTGCTGCTGCTGCTATGCTGCTGCTGCTGCT 484
139 Gly 139
      |||
485 GGT 487

seq_name: /SID58/gcdata/geneseq/geneseqn/NA2000.DAT:AAC33029

seq_documentation_block:
ID   AAC33029 standard; DNA; 1423 BP.
XX
AC   AAC33029;
XX
DT   17-OCT-2000 (first entry)
XX
DE   Arabidopsis thaliana DNA fragment SEQ ID NO: 1526.
XX
KW   Hybridisation assay; genetic mapping; gene expression control;
XX   protein identification; signal transduction pathway;
XX   metabolic pathway; promoter; termination sequence; ss.
XX
OS   Arabidopsis thaliana.
XX
PN   EP1033405-A2.
XX
PD   06-SEP-2000.
XX
PE   25-FEB-2000; 2000EP-0301439.
XX
PR   25-FEB-1999; 99US-0121825.
XX   05-MAR-1999; 99US-0123180.
XX   09-MAR-1999; 99US-0123548.
XX   23-MAR-1999; 99US-0125788.
XX   25-MAR-1999; 99US-0126264.
XX   29-MAR-1999; 99US-0126785.
XX   01-APR-1999; 99US-0127462.
XX   06-APR-1999; 99US-0128234.
XX   08-APR-1999; 99US-0128714.
XX   16-APR-1999; 99US-0129845.
XX   19-APR-1999; 99US-0130077.
XX   21-APR-1999; 99US-0130449.
XX   23-APR-1999; 99US-0130510.
XX   28-APR-1999; 99US-0130891.
XX   30-APR-1999; 99US-0131449.
XX   30-APR-1999; 99US-0132048.
XX   04-MAY-1999; 99US-0132407.
XX   05-MAY-1999; 99US-0132484.
XX   99US-0132485.
```

PR 06-MAY-1999;	99US-0132486.
PR 06-MAY-1999;	99US-0132487.
PR 07-MAY-1999;	99US-0132488.
PR 11-MAY-1999;	99US-0134256.
PR 14-MAY-1999;	99US-0134218.
PR 14-MAY-1999;	99US-0134219.
PR 14-MAY-1999;	99US-0134221.
PR 14-MAY-1999;	99US-0134370.
PR 18-MAY-1999;	99US-0134768.
PR 19-MAY-1999;	99US-0134941.
PR 20-MAY-1999;	99US-0135124.
PR 21-MAY-1999;	99US-0135353.
PR 24-MAY-1999;	99US-0135629.
PR 25-MAY-1999;	99US-0136021.
PR 27-MAY-1999;	99US-0136392.
PR 28-MAY-1999;	99US-0136782.
PR 01-JUN-1999;	99US-0137222.
PR 03-JUN-1999;	99US-0137528.
PR 04-JUN-1999;	99US-0137502.
PR 07-JUN-1999;	99US-0137724.
PR 08-JUN-1999;	99US-0138094.
PR 10-JUN-1999;	99US-0138540.
PR 10-JUN-1999;	99US-0138847.
PR 14-JUN-1999;	99US-0139119.
PR 16-JUN-1999;	99US-0139452.
PR 16-JUN-1999;	99US-0139453.
PR 17-JUN-1999;	99US-0139492.
PR 18-JUN-1999;	99US-0139454.
PR 18-JUN-1999;	99US-0139455.
PR 18-JUN-1999;	99US-0139456.
PR 18-JUN-1999;	99US-0139457.
PR 18-JUN-1999;	99US-0139458.
PR 18-JUN-1999;	99US-0139459.
PR 18-JUN-1999;	99US-0139460.
PR 18-JUN-1999;	99US-0139461.
PR 18-JUN-1999;	99US-0139462.
PR 18-JUN-1999;	99US-0139463.
PR 18-JUN-1999;	99US-0139750.
PR 18-JUN-1999;	99US-0139763.
PR 21-JUN-1999;	99US-0139817.
PR 22-JUN-1999;	99US-0139859.
PR 23-JUN-1999;	99US-0140353.
PR 23-JUN-1999;	99US-0140354.
PR 24-JUN-1999;	99US-0140695.
PR 28-JUN-1999;	99US-0140823.
PR 29-JUN-1999;	99US-0140991.
PR 30-JUN-1999;	99US-0141287.
PR 01-JUL-1999;	99US-0141842.
PR 01-JUL-1999;	99US-0142154.
PR 02-JUL-1999;	99US-0142055.
PR 06-JUL-1999;	99US-0142390.
PR 08-JUL-1999;	99US-0142803.
PR 09-JUL-1999;	99US-0142920.
PR 12-JUL-1999;	99US-0142977.
PR 13-JUL-1999;	99US-0143542.
PR 14-JUL-1999;	99US-0143624.
PR 15-JUL-1999;	99US-0144005.
PR 16-JUL-1999;	99US-0144085.
PR 16-JUL-1999;	99US-0144086.
PR 19-JUL-1999;	99US-0144325.
PR 19-JUL-1999;	99US-0144331.
PR 19-JUL-1999;	99US-0144332.
PR 19-JUL-1999;	99US-0144333.
PR 19-JUL-1999;	99US-0144334.
PR 19-JUL-1999;	99US-0144335.
PR 20-JUL-1999;	99US-0144352.
PR 20-JUL-1999;	99US-0144632.
PR 20-JUL-1999;	99US-0144884.
PR 21-JUL-1999;	99US-0144814.
PR 21-JUL-1999;	99US-0145086.
PR 21-JUL-1999;	99US-0145088.
PR 22-JUL-1999;	99US-0145085.
PR 22-JUL-1999;	99US-0145087.

PR 22-JUL-1999;	99US-0145089.
PR 22-JUL-1999;	99US-0145192.
PR 23-JUL-1999;	99US-0145145.
PR 23-JUL-1999;	99US-0145218.
PR 23-JUL-1999;	99US-0145224.
PR 26-JUL-1999;	99US-0145276.
PR 27-JUL-1999;	99US-0145913.
PR 27-JUL-1999;	99US-0145918.
PR 27-JUL-1999;	99US-0145919.
PR 28-JUL-1999;	99US-0145951.
PR 02-AUG-1999;	99US-0146386.
PR 02-AUG-1999;	99US-0146388.
PR 02-AUG-1999;	99US-0146389.
PR 03-AUG-1999;	99US-0147038.
PR 04-AUG-1999;	99US-0147204.
PR 04-AUG-1999;	99US-0147302.
PR 05-AUG-1999;	99US-0147192.
PR 05-AUG-1999;	99US-0147260.
PR 06-AUG-1999;	99US-0147303.
PR 06-AUG-1999;	99US-0147416.
PR 09-AUG-1999;	99US-0147493.
PR 09-AUG-1999;	99US-0147935.
PR 10-AUG-1999;	99US-0148171.
PR 11-AUG-1999;	99US-0148319.
PR 12-AUG-1999;	99US-0148341.
PR 13-AUG-1999;	99US-0148565.
PR 13-AUG-1999;	99US-0148684.
PR 16-AUG-1999;	99US-0149368.
PR 17-AUG-1999;	99US-0149175.
PR 18-AUG-1999;	99US-0149426.
PR 20-AUG-1999;	99US-0149722.
PR 20-AUG-1999;	99US-0149723.
PR 20-AUG-1999;	99US-0149929.
PR 23-AUG-1999;	99US-0149902.
PR 23-AUG-1999;	99US-0149930.
PR 25-AUG-1999;	99US-0150566.
PR 26-AUG-1999;	99US-0150884.
PR 27-AUG-1999;	99US-0151065.
PR 27-AUG-1999;	99US-0151066.
PR 30-AUG-1999;	99US-0151080.
PR 30-AUG-1999;	99US-0151302.
PR 31-AUG-1999;	99US-0151438.
PR 01-SEP-1999;	99US-0151930.
PR 07-SEP-1999;	99US-0152363.
PR 10-SEP-1999;	99US-0153070.
PR 13-SEP-1999;	99US-0153758.
PR 15-SEP-1999;	99US-0154018.
PR 16-SEP-1999;	99US-0154039.
PR 20-SEP-1999;	99US-0154779.
PR 22-SEP-1999;	99US-0155139.
PR 23-SEP-1999;	99US-0155486.
PR 24-SEP-1999;	99US-0155659.
PR 28-SEP-1999;	99US-0156458.
PR 29-SEP-1999;	99US-0156596.
PR 04-OCT-1999;	99US-0157117.
PR 05-OCT-1999;	99US-0157753.
PR 06-OCT-1999;	99US-0157865.
PR 07-OCT-1999;	99US-0158029.
PR 08-OCT-1999;	99US-0158232.
PR 12-OCT-1999;	99US-0158369.
PR 13-OCT-1999;	99US-0159283.
PR 13-OCT-1999;	99US-0159294.
PR 13-OCT-1999;	99US-0159295.
PR 14-OCT-1999;	99US-0159329.
PR 14-OCT-1999;	99US-0159330.
PR 14-OCT-1999;	99US-0159331.
PR 14-OCT-1999;	99US-0159637.
PR 14-OCT-1999;	99US-0159638.
PR 18-OCT-1999;	99US-0159584.
PR 21-OCT-1999;	99US-0160741.
PR 21-OCT-1999;	99US-0160767.
PR 21-OCT-1999;	99US-0160768.
PR 21-OCT-1999;	99US-0160770.

PR 21-OCT-1999; 99US-0160814.
 PR 21-OCT-1999; 99US-0160815.
 PR 22-OCT-1999; 99US-0160880.
 PR 22-OCT-1999; 99US-0160981.
 PR 22-OCT-1999; 99US-0160989.
 PR 23-OCT-1999; 99US-0161404.
 PR 23-OCT-1999; 99US-0161405.
 PR 25-OCT-1999; 99US-0161406.
 PR 26-OCT-1999; 99US-0161359.
 PR 26-OCT-1999; 99US-0161360.
 PR 26-OCT-1999; 99US-0161361.
 PR 28-OCT-1999; 99US-0161920.
 PR 28-OCT-1999; 99US-0161992.
 PR 28-OCT-1999; 99US-0161993.
 PR 29-OCT-1999; 99US-0162142.

alignment_scores:
 Quality: 180.50 Length: 141
 Ratio: 2.375 Gaps: 11
 Percent Similarity: 53.901 Percent Identity: 41.844

alignment_block:
 US-09-554-547-16 x AAC33029 ..

Align seg 1/1 to: AAC33029 from: 1 to: 1423

33 ArgValHisGly.....AsnLeuTyrAlaGlyIleGluArgAlaG1 46
 ||| |||||
 793 AGGTCAATGTGAGAGTCTCTAGTATGCTGCG..... 828
 46 yProArGlyTyrProGlyLeuThrAlaSerIleGlyGluValG1A 63
 |||||
 829GGATATGGGTATGCTGCTGCACTAGAGCCCTGAATTGGA. 870
 63 laArgLeuGlyArgAlaGlyValGlyAlaSerTyrGlyGly 79
 |||||
 871GGAGCTTATATATTAATTAATGAGGGGTGAGTTAGA...GGT 909
 80 Tyr.....ProSerTyrGlyTyr.....ProTyrG1 88
 ||| |||
 910 TACCGAATGAACACACCCCTGGCTATTCATGCTTTTGGACCCATATG 959
 88 y.....GlyTyrGly.....GlyTyrGlyGlyTyrG 100
 |||||
 960 GAGTGGCTTTGGTGTGAGGATATGCGCGTGAAGAGAGACCATTT 1009
 100 lylGlyTyrAspGlnGlyPheGlySerAlaTyrGlyGlyTyr.....Pro 114
 |||||
 1010 TGGGATATCTCTGCTGAGAGAGGGAAGGCTATGGGGATATGTCGGCTT 1059
 115 GlyTyrTyrGlyTyrTyrTyrProSerGlyTyrGlyGlyGlyTyr..... 129
 ||||| ||| |||
 1060 GGTTCAGCGCGGTGATATGATCTGCTGCGCCAGTGTGAGCATATGAGG 1109
 130GlyGlySerTyrGlyGlySerTyrGlyGlySerTyrThTyr...P 144
 ||||| ||||| |||||
 1110 AGCAGGCTGCTCATATGAGGAGGCTTATAGTACAGTACATACATC 1159
 144 roAsnValArgAlaSerAlaGly 151
 |||
 1160 CGTATGCAGAGTGAAGGTCAGAGT 1182

seq_name: /SID8/gcdata/geneseq/geneseqn/NA2000.DAT:AAF18282

seq_documentation_block:

ID AAF18282 standard; DNA; 1089 BP.

AC AAF18282;

14-MAR-2001 (first entry)

DE Lung cancer associated polynucleotide sequence SEQ ID 301.

KW Human; lung cancer associated protein; neuroprotective; cyostatic;
 KW cardioactive; immunomodulatory; muscular active; vlnetary;
 KW gastrointestinal; nephrotoxic; antiinfective; gynecological;
 KW antibacterial; diagnosis; neural disorder; immune disorder; reproductive;
 KW proliferative disorder; wound healing; infectious disease; ds.
 XX
 OS Homo sapiens.
 XX
 PN WO20005180-A2.
 XX
 PD 21-SEP-2000.
 XX
 PF 08-MAR-2000; 2000WO-US05918.
 XX
 PR 12-MAR-1999; 99US-0124270.
 XX
 PA (HUMA-) HUMAN GENOME SCT INC.
 PA (ROSE/) ROSEN C A.
 XX
 PI Ruben SM;
 XX
 DR WPI: 2000-587514/55.
 DR P-PSDB: AAB58406.
 XX
 PT Lung cancer associated gene sequences, referred to as lung cancer
 PT antigens, useful for treatment, prevention, and diagnosis of disorders
 PT such as lung cancer -
 XX
 PS Claim 1; Page 759; 1425pp; English.

XX Polynucleotide sequences AAF17982 - AAF18424 encode human lung cancer
 CC associated proteins represented in AAB58106 - AAB58548. Lung cancer
 CC associated proteins and polynucleotide sequences, their agonists, and
 CC antagonists may have neuroprotective; cyostatic; cardioactive;
 CC immunomodulatory; muscular active general; vulnerary; gastrointestinal
 CC general; nephrotoxic; antiinfective; gynecological; or antibacterial
 CC activity. The invention also includes antibodies specific for the
 CC protein or polynucleotide sequences. The lung cancer associated
 CC polynucleotide sequences may be used for detection of lung cancer,
 CC chromosome identification, as chromosome markers, and for numerous other
 CC diagnostic or research purposes. The proteins may be used to treat
 CC disorders such as neural, immune, muscular, reproductive,
 CC gastrointestinal, pulmonary, cardiovascular, renal, and proliferative
 CC disorders. The proteins may also be used in the treatment of wounds and
 CC infectious diseases. Polynucleotide sequences AAF18425 - AAF18433 and
 CC peptide AAB58549 are used in the course of the invention for the
 CC identification and characterisation of the polynucleotide and protein
 CC sequences.
 XX
 SQ Sequence 1089 BP; 289 A; 183 C; 294 G; 316 T; 7 other;

alignment_scores:
 Quality: 177.50 Length: 148
 Ratio: 2.219 Gaps: 10
 Percent Similarity: 54.054 Percent Identity: 40.541

alignment_block:
 US-09-554-547-16 x AAF18282 ..

Align seg 1/1 to: AAF18282 from: 1 to: 1089

14 AlaLeuThrAsnAlaAlaArgAlaGlyArgLeuGlySerAspLeuAspTh 30
 ||||| ||| |||||
 54 GCATTAGCAATTTCCCAAGGCACTCTTTAAGCGCAGCTGCTCTCTATC 103
 30 rPheGlyArgValHisGlyAsn.....LeuTyrAla...G 41
 ||||| |||||
 104 CCTTGA.....GGCATCAGTAAATGTTTCCATCATATAGCAGAG 144
 41 lylleGluArgAlaGlyPProArGlyTyrProGlyLeuThrAlaSerIle 57
 ||| ||||| |||||
 145 GTAGTTATGAGAGGATGATGATATATATGGA.....TTT 182

XX Arabidopsis thaliana.
OS
XX
PN EPI03405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
XX 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0128845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 28-APR-1999; 99US-0130891.
PR 30-APR-1999; 99US-0131449.
PR 04-MAY-1999; 99US-0132407.
PR 05-MAY-1999; 99US-0132484.
PR 06-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 07-MAY-1999; 99US-0132487.
PR 11-MAY-1999; 99US-0132563.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 18-MAY-1999; 99US-0134370.
PR 19-MAY-1999; 99US-0134768.
PR 20-MAY-1999; 99US-0134941.
PR 21-MAY-1999; 99US-0135124.
PR 24-MAY-1999; 99US-0135353.
PR 25-MAY-1999; 99US-0135629.
PR 27-MAY-1999; 99US-0136021.
PR 28-MAY-1999; 99US-0136392.
PR 01-JUN-1999; 99US-0136782.
PR 03-JUN-1999; 99US-0137222.
PR 04-JUN-1999; 99US-0137528.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139494.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.

PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 22-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 26-JUL-1999; 99US-0145324.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149930.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.


```

PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158359.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160757.
PR 21-OCT-1999; 99US-0160758.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161044.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 26-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

```

alignment_scores:

```

Quality: 175.50 Length: 147
Ratio: 2.372 Gaps: 6
Percent Similarity: 50.340 Percent Identity: 36.735

```

alignment_block:

```
US-09-554-547-16 x AAC44994 ..
```

```
Align seg 1/1 to: AAC44994 from: 1 to: 993
```

```

2 LysAlaPhePheValLeuSerThrAlaLeuThrAsnAl 18
   :::::||||| :::::||||| :::::|||||
304 AGCGATTTGGATTGTTTACATTCACCTTTCAGAACGCCGCTACGTGC 353
   18 aAlaArgAla.....GlyArgLeuGlySerAspLeuAspThrPheGlyA 33
      I :::::|||||
      354 TATCCACGCGCTTTCGATGCCGCCG.....GATTCACCATGCCGC 388
      |||||
339 GGGTCTTAAACTAAACTATGCAAAATGATAGAACAAAGTGT..... 429
      |||||
33 rGValHISGlyAsnLeuTyArgAlaGlyIleGlyAlaArgAlaGlyProArgGly 49
      |||||
50 TyrProGlyLeuThrAlaSerIleGlyGlyValGlyAlaArgLeuGly 66
      |||||
430 .....GGTGGTTTGGAGGTGGTGGCTATG 455
      |||||
66 yGlyArgAlaGlyValGlyValSerSerTyArgIleTyArgIleTyProSerT 83
      |||||
456 AGCGCGTGTGTGTGTATGCTGTAGTGTGCGTACGCA..... 495

```

```

seq_name: /SID58/gcdata/geneseq/geneseqn/NA2001.DAT:AA158656
seq_documentation_block:
ID AA158656 standard; cDNA: 1505 BP.
XX
AC AA158656;
XX
DT 22-OCT-2001 (first entry)
XX
DE Human polynucleotide SEQ ID NO 859.
XX
KW Human; nootropic; immunosuppressant; cytosstatic; gene therapy; cancer;
KW peripheral nervous system; neuropathy; central nervous system; CNS;
KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
KW chemokine; thrombolytic; drug screening; arthritis; inflammation;
KW leukaemia; ss.
XX
OS Homo sapiens.
XX
PN M0200153312-A1.
XX
PD 26-JUL-2001.
XX
PF 26-DEC-2000; 2000MO-US34263.
XX
PR 21-JAN-2000; 2000US-0488725.
PR 25-APR-2000; 2000US-0552317.
PR 09-JUL-2000; 2000US-0598042.
PR 19-JUL-2000; 2000US-0620312.
PR 03-AUG-2000; 2000US-0653450.
PR 14-SEP-2000; 2000US-0662191.
PR 19-OCT-2000; 2000US-0693036.
PR 29-NOV-2000; 2000US-0727344.
XX
XX (HYSE-) HYSEQ INC.
XX
XX Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
XX Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;
XX Zhao QA, Zhou P, Goodrich R, Drmanac RT;
XX WPT: 2001-442253/47.
XX P-PSDB: AAM39500.
XX
XX Novel nucleic acids and polypeptides, useful for treating disorders
XX such as central nervous system injuries -
XX
XX Claim 1: SEQ ID NO 859; 10078bp; English.
XX
XX The invention relates to human nucleic acids (AA157798-AA161369) and
XX the encoded polypeptides (AAM38642-AA42213) with nootropic,
XX immunosuppressant and cytosstatic activity. The polynucleotides are useful
XX in gene therapy. A composition containing a polypeptide or polynucleotide
XX of the invention may be used to treat diseases of the peripheral nervous
XX system, such as peripheral nervous injuries, peripheral neuropathy and
XX localised neuropathies and central nervous system diseases, such as

```



```

2266 GGTGGAGGATACAAAGCCCATCACCGTCGCGGCGTACGAGGTCGTGGACA 2315
91 .....GlyGlyTyrGlyGlyTyrGlyGlyTyr 99
2316 AGGAGCGCCGTGATGTGATCTGCGAGGAGGATATGGAGCGCAATGGTA 2365
99 YrGlyGlyTyr.....AspGlnGlyHegGlySerAlaTyrGlyGlyTyr 113
||||| |||:||||| |||||:|||||
2366 GTTCCTGGGTACGGCGCATGACATGATGCTGTGTGCTGCTGCTGCGGGGCA 2415
114 ProGlyTyrTyrGlyTyrTyrTyrTyrProSerGlyTyrGlyGlyGlyTyrGly 130
||||| ||| ||||| ||||| ||||| |||||
2416 AATGCGCATGCTGCTG.....GGTACGAGAGAGAGAGAGAG 2450
130 YGlySerTyrGlyGlySerTyrGly...GlySerTyrThrTyrProAsnY 146
||||:|||||:||||| ||||| ||||| |||||
2451 CGCGCGTATATGAGCGCGCTACGGCAGCTGTAGTGTACAGCAGCAGAG 2500
146 aLArgAlaSerAlaGlyAlaAlaAla 154
:::|||||:|||||
2501 GCGGATATCATGGCGCGCTGTGTCA 2526

seq_name: /SID58/gcgdata/geneseq/geneseqn/NA2001.DAT:AAI60442
seq_documentation_block:
ID AAI60442 standard; cDNA: 3678 BP.
XX
AC AAI60442:
XX
DT 22-OCT-2001 (first entry)
XX
DE Human polynucleotide SEQ ID NO 4431.
XX
Human; neotropic; immunosuppressant; cytostatic; gene therapy; cancer;
peripheral nervous system; neuropathy; central nervous system; CNS;
Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
amniotic; lateral sclerosis; Shy-Drager Syndrome; chemotactic;
chemokine; thrombolytic; drug screening; arthritis; inflammation;
leukaemia; ss.
XX
OS Homo sapiens.
XX
PN W020015312-A1.
XX
PD 26-JUL-2001.
XX
PF 26-DEC-2000; 2000MO-US34263.
XX
PR 25-JAN-2000; 2000US-0488725.
PR 25-APR-2000; 2000US-0552317.
PR 09-JUL-2000; 2000US-0598042.
PR 19-AUG-2000; 2000US-0620312.
PR 03-AUG-2000; 2000US-0653450.
PR 14-SEP-2000; 2000US-0662191.
PR 19-OCT-2000; 2000US-0693036.
PR 29-NOV-2000; 2000US-0727344.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D,
PI Wang J, Wang Z, Weinman T, Xu C, Xue AJ, Yang Y, Zhang J;
PI Zhao QX, Zhou P, Goodrich R, Drmanec RT;
XX
XX WPI: 2001-442253/47.
XX P-PSDB; AAM41266.
XX
XX Novel nucleic acids and polypeptides, useful for treating disorders
XX such as central nervous system injuries -
XX
XX Claim 1; SEQ ID NO 4431; 10078bp; English.
XX
XX The invention relates to human nucleic acids (AAI57798-AAI61369) and
XX the encoded polypeptides (AAM38642-AAM42213) with neotropic,

```

[illegible]

[illegible]

```

139 ySerTyrThrTyrProAsnValArgAlaSerAlaIleAlaAla 154
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
2257 ACAGAGACATAT...GGTCCGAGCTGGAGCGGTGCTGCTGCA 2299

seq_name: /cgn2_6/plodata/2/1na/6A_COMB.seq:US-08-963-168C-3

seq_documentation_block:
: Sequence 3, Application US/08963168C
: Patent No. 6127166
: GENERAL INFORMATION:
: APPLICANT: Bayley, Hagan
: APPLICANT: Cao, Quidjan
: APPLICANT: Wang, Yunjaun
: TITLE OF INVENTION: MOLLUSCAN LIGAMENT POLYPEPTIDES
: NUMBER OF SEQUENCES: 43
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Fish & Richardson P.C.
: STREET: 225 Franklin Street
: CITY: Boston
: STATE: MA
: COUNTRY: US
: ZIP: 02110-2804
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Diskette
: OPERATING SYSTEM: Windows95
: SOFTWARE: FastSeq for Windows Version 2.0
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/963.168C
: FILING DATE: 03-NOV-1997
: ATTORNEY/AGENT INFORMATION:
: NAME: Fasse, Peter J.
: REGISTRATION NUMBER: 32,983
: REFERENCE/DOCKET NUMBER: 07917/059001
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 617/542-5070
: TELEFAX: 617/542-8906
: TELEX: 200154
: INFORMATION FOR SEQ ID NO: 3:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 381 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: cDNA
: FEATURE:
: NAME/KEY: Coding Sequence
: LOCATION: 1...378
: US-08-963-168C-3

Alignment_scores:
      Quality: 171.00      Length: 140
      Ratio: 2.085      Gaps: 7
      Percent Similarity: 58.571      Percent Identity: 36.429

Alignment_block:
US-09-554-547-16 x US-08-963-168C-3  ..

Align seg 1/1 to: US-08-963-168C-3 from: 1 to: 381

1 MetLysAlaPhePheValLeuSerLeuLeuSerThrAlaAlaLeuThrAs 17
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
1 ATGAACGCTTACATCTGCTTTCTGCTTCTTATTCGCTGCTGCAAGCC 50
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
17 nAlaAlaLeuArgAlaGlyArgLeuGlySerAspLeuAspThrPheGlyArgV 34
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
51 CCGCGGATGAGGAGGTGCTGCGGAGT...ATGGCGCGTACCGGAGAA 97
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
34 wHisGlyAsnLeuTyrAlaGlyIleGlyAlaArgAlaGlyProArgGlyTyr 50
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
98 TGGGAGGCGGAGTGAACGAGGC.....GGATTTC 126

```

```

51 ProGlyLeuThrAlaSerIleGlyGlyValGlyAlaArgLeuGlyG 67
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
127 GCGGATATGCGCGAGGAATGGCGAGGTAAAGCGGA...TTCGGCGG 173
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
67 yArgAlaGlyValGlyValSerSerTyrGlyTyrGlyTyrProSerTrpG 84
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
174 AATGGCGGATTCGGCGCATGGAGGTGAATGGGC.....G 211
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
84 lYtYrProTyrGlyGlyTyrGlyGlyTyrGlyGlyTyrGlyGly 100
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
212 GAGGTCA...GGCGATTCGTTGGAATGGAGGTTTCGAGGAATGGCT 258
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
101 GlyTyrAspGlnGlyPheGlySerAlaTyrGlyGlyTyrProGlyTyrTy 117
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
259 GCGGAAAGGTGATTCGAGGAATGGCAGTGTGATGGAGGTTTCG 308
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
117 rGlyTyrTyrTyrProSerGlyTyrGlyGlyGlyTyrGlyGlySerTyG 134
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
309 AGGAATG.....GGAGCGGAATGCCGTTTCGGTGAATGGCGC 349
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
134 lY...GlySerTyrGlyGly 139
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
350 GCCAAGTGGATTTCGCGCA 369

seq_name: /cgn2_6/plodata/2/1na/6A_COMB.seq:US-08-963-168C-1

seq_documentation_block:
: Sequence 1, Application US/08963168C
: Patent No. 6127166
: GENERAL INFORMATION:
: APPLICANT: Bayley, Hagan
: APPLICANT: Cao, Quidjan
: APPLICANT: Wang, Yunjaun
: TITLE OF INVENTION: MOLLUSCAN LIGAMENT POLYPEPTIDES
: NUMBER OF SEQUENCES: 43
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Fish & Richardson P.C.
: STREET: 225 Franklin Street
: CITY: Boston
: STATE: MA
: COUNTRY: US
: ZIP: 02110-2804
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Diskette
: OPERATING SYSTEM: Windows95
: SOFTWARE: FastSeq for Windows Version 2.0
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/963.168C
: FILING DATE: 03-NOV-1997
: ATTORNEY/AGENT INFORMATION:
: NAME: Fasse, Peter J.
: REGISTRATION NUMBER: 32,983
: REFERENCE/DOCKET NUMBER: 07917/059001
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 617/542-5070
: TELEFAX: 617/542-8906
: TELEX: 200154
: INFORMATION FOR SEQ ID NO: 1:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 411 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: cDNA
: FEATURE:
: NAME/KEY: Coding Sequence
: LOCATION: 1...408
: OTHER INFORMATION: The amino acid translation for SEQ ID NO. 6127166. 1
: US-08-963-168C-1

```

```
; CURRENT APPLICATION DATA:
```



```
seq_documentation_block:
; Sequence 1, Application US/07915246
; Patent No. 5401836
; GENERAL INFORMATION:
; APPLICANT: Basczynski, Chris L.
; APPLICANT: Falls, Lynne
; APPLICANT: Bellmare, Guy
; APPLICANT: Boivin, Rodolphe
; TITLE OF INVENTION: A BRASSICA REGULATORY SEQUENCE FOR
; TITLE OF INVENTION: ROOT-SPECIFIC OR ROOT-ABUNDANT GENE EXPRESSION
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: McAndrews, Held, and Malloy
; STREET: 500 W. Madison St. 34th Floor
; CITY: Chicago
; STATE: Illinois
; COUNTRY: USA
; ZIP: 60661
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/915,246
; FILING DATE: 19920716
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Pochopien, Donald J.
; REGISTRATION NUMBER: 32,167
; REFERENCE/DOCKET NUMBER: 91 P 1125
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312 707-8889
; TELEFAX: 312 707-9155
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1505 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHEtical: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Brassica napus
; STRAIN: Westar
; DEVELOPMENTAL STAGE: Somatic
; TISSUE TYPE: Root
; US-07-915-246-1

alignment_scores:
; Quality: 169.00 Length: 161
; Ratio: 2.036 Gaps: 6
; Percent Similarity: 51.553 Percent Identity: 36.646

alignment_block:
US-09-554-547-16 x US-07-915-246-1 ..
Align seg 1/1 to: US-07-915-246-1 from: 1 to: 1505

2 lysalaphelvalleuserleuserthrallalaleuthasnal 18
||| :||| ||| :||| :||| :||| :||| :|||
564 AAAATCTTTCTTCCTATGCTTCTTATAGCTTAAAGGTTATGTTCTGC 613
18 aalargalaglyargleuglyseraspleuaspthrphelyargvalh 35
||| ||| ||| ||| ||| ||| ||| |||
614 AAGACGACGA.....CTTCTTCTCTCCATGACCCGACGATGAAGTCG 657
35 lsgly.....asleutyralaglyle..... 42
658 CCGGATACGGCGAAGAAAGTAGTTCGATGCTGTTATGATGATTGAGCTT 707
```

```
43 .....gluargalaglyproargl 49
708 GATGCTGCTGTGCTGTGAGAGCTGCTGCGGAGAGAGCTGCTGCTG 757
49 ytyrproglyleuthrallaserileglyglvalglaaargleug 66
||| ||| :||| :||| :||| :||| :||| :|||
758 TTACGCTGA...GCTGAAGCATTGCTGAGAGAGAGCGCTGACATG 804
66 lylalargalaglyalaglyalsersertyrlytyrlytyrproser 82
||| ||| ||| ||| ||| ||| ||| |||
805 GTGCTGCTGCTGCTGAGCT.....GCTGCTGCTGCTGCTGCTGA 842
83 trpilytyrprotyrlytyrlytyrlytyrlytyrlytyrlytyr 99
||| ||| ||| ||| ||| ||| ||| |||
843 GGAGATCTGCTTATGAGCTGGAAGCGGTGAAGCTGCTGCTGATA 892
99 rglylytyraspgln.....glyheglyseralatyrglyg 112
||| ||| ||| ||| ||| ||| ||| |||
893 CCGAGCGGAGAGCTGAGAGACATGCTGAGTGAAGAGCGAGAG 942
112 ltyrproglytyrtyrlytyrtyrtyrtyrtyrtyrtyrtyrtyr 128
||| ||| ||| ||| ||| ||| ||| |||
943 GTGCTGCTGAGAGACTGCGCGTCCCATGCTGCTGATGATGCTGCTGA 992
129 tyrglylyserlytyrlytyrlytyrlytyrlytyrlytyrly 139
||| ||| ||| ||| ||| ||| ||| |||
993 GAAGCTGCTGCTGCTGAGAGAGATATGAGCT 1025

seq_name: /cgn2_6/ptodata/2/ina/6A_COMB.seq:US-08-963-168C-5

seq_documentation_block:
; Sequence 5, Application US/08963168C
; Patent No. 6127166
; GENERAL INFORMATION:
; APPLICANT: Bayley, Hagan
; APPLICANT: Cao, Qunping
; APPLICANT: Wang, Yunjaun
; TITLE OF INVENTION: MOLUSCAN LIGAMENT POLYPEPTIDES
; TITLE OF INVENTION: AND GENES ENCODING THEM
; NUMBER OF SEQUENCES: 43
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: MA
; COUNTRY: US
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM compatible
; OPERATING SYSTEM: Windows95
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/963,168C
; FILING DATE: 03-NOV-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Fasse, Peter J.
; REGISTRATION NUMBER: 32,983
; REFERENCE/DOCKET NUMBER: 07917/059001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617/542-5070
; TELEFAX: 617/542-8906
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 399 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: Coding Sequence
; LOCATION: 1...396
```

SOFTWARE: FASTSEQ for Windows Version 2.0

```
seq_documentation_block:
```

```
Sequence 1, Application US/0814777
Patent No. 5914265
GENERAL INFORMATION:
APPLICANT: Rood, Dennis R.
APPLICANT: Rothnagel, Joseph A.
APPLICANT: Greenhalgh, David A.
APPLICANT: Yuspa, Stuart H.
TITLE OF INVENTION: KERATIN K1 EXPRESSION VECTORS
TITLE OF INVENTION: AND METHODS OF USE
NUMBER OF SEQUENCES: 5
CORRESPONDENCE ADDRESS:
ADDRESSEE: LYON & LYON
STREET: 611 West Sixth Street
CITY: Los Angeles
STATE: California
COUNTRY: U.S.A.
ZIP: 90017
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5" Diskette, 1.44 Mb storage
COMPUTER: IBM PC compatible
OPERATING SYSTEM: IBM MS-DOS (Version 5.0)
SOFTWARE: Wordperfect (Version 5.1)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/147,777
FILING DATE:
CLASSIFICATION: 800
PRIOR APPLICATION DATA:
PRIOR APPLICATION DATA: Including application
PRIOR APPLICATION DATA: described below:
APPLICATION NUMBER: 07/876,289
FILING DATE: April 30, 1992
APPLICATION NUMBER: Unassigned (204/144)
FILING DATE: October 29, 1993
ATTORNEY/AGENT INFORMATION:
NAME: Warburg, Richard J.
REGISTRATION NUMBER: 32,327
REFERENCE/DOCKET NUMBER: 204/153
TELECOMMUNICATION INFORMATION:
TELEPHONE: (213) 489-1600
TELEFAX: (213) 955-0440
TELEX: 67-3510
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 10747 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-147-777-1

Alignment_scores:
Quality: 156.50 Length: 101
Ratio: 2.524 Gaps: 3
Percent Similarity: 61.386 Percent Identity: 43.564

alignment_block:
US-09-554-547-16 x US-08-147-777-1 ..

Align seg 1/1 to: US-08-147-777-1 from: 1 to: 10747

54 ThrAlaSerIleGlyGlyValGlyAlaArgLeuGlyAlaArgAlaG1 70
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
6056 ACCATCATGTGGAGTGGCGGCGGAGGAGGAGTGGC...GGCGGTGGCTACGG 6102
70 yValGlyValSerSerTyGlyTyGlyTyGlyProSerTrpGlyTyPProT 87
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
6103 CTCTGAGAGTACACAGCTATAGCTCCGAGAGGTGATGCTATGCTTCGAGG 6152
87 yGlyGlyTyGlyTyGlyTyGlyTyGlyTyGlyTyGlyTyGlyTyGlyTyGly 103
||||| |||||| |||||| |||||| |||||| |||||| |||||| |||
6153 GTGGCGGCGCGCGCGCGGTGGCGAGCTATGGCTCC...GAGAGTGGAGC 6199
```

```
104 GInglyPheGlySerAlaTyGlyGlyTyGlyTyGlyTyGlyTyGly 120
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
6200 TATGGCTGTGAGGTGGCGGCGGCGCATGTGCAGCTACGCTCCGAGG 6249
120 rTyPProSerGlyTyGlyGlyTyGlyTyGlyTyGlyTyGlyTyGly 137
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
6250 CAGCAGTGGGGGCTACAGAGGTGGCTGTGGAGCGCGGCGGCGCAGCT 6299
137 yGlyGlySerTyTrpTrpProAsnValArgAlaSerAlaGlyAla 153
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
6300 CTGGCGGC.....CGGCGCTCTGGCGGCGGAGC 6328
154 Ala 154
|||
6329 TCT 6331

seq_name: /cgn2_6/ptodata/2/ina/6A_COMB.seq:US-08-452-872-1
seq_documentation_block:
Sequence 1, Application US/08452872
Patent No. 6057298
GENERAL INFORMATION:
APPLICANT: Rood, Dennis R.
APPLICANT: Rothnagel, Joseph A.
APPLICANT: Greenhalgh, David A.
APPLICANT: Yuspa, Stuart H.
TITLE OF INVENTION: KERATIN K1 EXPRESSION VECTORS
TITLE OF INVENTION: AND METHODS OF USE
NUMBER OF SEQUENCES: 5
CORRESPONDENCE ADDRESS:
ADDRESSEE: LYON & LYON
STREET: 611 West Sixth Street
CITY: Los Angeles
STATE: California
COUNTRY: U.S.A.
ZIP: 90017
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5" Diskette, 1.44 Mb storage
COMPUTER: IBM PC compatible
OPERATING SYSTEM: IBM MS-DOS (Version 5.0)
SOFTWARE: Wordperfect (Version 5.1)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/452,872
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/147,777
FILING DATE:
APPLICATION NUMBER: 07/876,289
FILING DATE: April 30, 1992
APPLICATION NUMBER: Unassigned (204/144)
FILING DATE: October 29, 1993
ATTORNEY/AGENT INFORMATION:
NAME: Warburg, Richard J.
REGISTRATION NUMBER: 32,327
REFERENCE/DOCKET NUMBER: 204/153
TELECOMMUNICATION INFORMATION:
TELEPHONE: (213) 489-1600
TELEFAX: (213) 955-0440
TELEX: 67-3510
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 10747 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-452-872-1

alignment_scores:
Quality: 156.50 Length: 101
Ratio: 2.524 Gaps: 3
```

Percent Similarity: 61.386 Percent Identity: 43.564

alignment_block:
US-09-554-547-16 x US-08-452-872-1 ..

Align seg 1/1 to: US-08-452-872-1 from: 1 to: 10747

```

54 ThrAlaSerLleGlyGlyGlyAlaValArgLeuGlyGlyArgAlaG1 70
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
6056 ACCATCAGTGGAGGTGGCAGCCAGGAGGTGGC...GGCGTGGCTACGG 6102
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
70 yValGlyValSerSerTyrglyTyrglyTyrrProSerTTPGlyTyrrProT 87
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
6103 CTCTGGAGGTAGCAGCTATGCTCCGAGGTGCTAGCTATGTTCTGGAG 6152
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
87 yrclyGlyTyrglyTyrglyTyrglyTyrglyTyrrProSerTTPGlyTyrrProT 103
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
6153 GTGGCGGGCGGGCGCGCGCTGCGAGCTATGCTCC...GGAGGTGGCAGC 6199
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
104 GlnGlyPheGlySerAlaTyrglyGlyTyrrProGlyTyrrTyrglyTyrrTy 120
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
6200 TATGGCTCTGGAGGTGGCGGGCGGCGCATGGCAGCTACCGCTCCGGAAG 6249
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
120 rTyrrProSerGlyTyrglyGlyTyrglyTyrglyTyrrTyrrTyrrTyrr 137
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
6250 CACACAGTGGGGCTACAGAGGTGCTCTGGAGGCGGGCGGCGGCGAGCT 6299
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
137 yrclyGlySerTyrrTyrrProAsnValArgAlaSerAlaGlyAlaAla 153
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
6300 CTGGCGGC.....CGGGCTCTGGCGGGGAGC 6328
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
154 Ala 154
6329 TCT 6331

```

seq_name: /cgn2_6/ptodata/2/1na/PCTUS_COMB.seq: PCT-US93-03985-1

seq_documentation_block:

Sequence 1, Application PC/TUS9303985

GENERAL INFORMATION:

APPLICANT: Roop, Dennis R.
APPLICANT: Rothnagel, Joseph A.
APPLICANT: Greenhalgh, David A.
APPLICANT: Yuspa, Stuart H.
TITLE OF INVENTION: DEVELOPMENT OF A VECTOR TO TARGET GENE
TITLE OF INVENTION: EXPRESSION TO THE EPIDERMIS OF TRANSGENIC ANIMALS
NUMBER OF SEQUENCES: 5
CORRESPONDENCE ADDRESS:
ADDRESSEE: Fulbright & Jaworski
STREET: 1301 McKinney, Suite 5100
CITY: Houston
STATE: Texas
COUNTRY: U.S.A.
ZIP: 77010-3095

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: PCT/US93/03985

FILING DATE: 19930428

CLASSIFICATION:

ATTORNEY/AGENT INFORMATION:

NAME: Paul, Thomas D. 32,714

REGISTRATION NUMBER: D-5478

TELECOMMUNICATION INFORMATION:

TELEPHONE: 713/651-5325

TELEFAX: 713/651-5246

TELEX: 762829

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:

```

; LENGTH: 10747 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; PCT-US93-03985-1

```

alignment_scores:

Quality: 156.50 Length: 101
Ratio: 2.524 Gaps: 3
Percent Similarity: 61.386 Percent Identity: 43.564

alignment_block:

US-09-554-547-16 x PCT-US93-03985-1 ..

Align seg 1/1 to: PCT-US93-03985-1 from: 1 to: 10747

```

54 ThrAlaSerLleGlyGlyGlyAlaValArgLeuGlyGlyArgAlaG1 70
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
6056 ACCATCAGTGGAGGTGGCAGCCAGGAGGTGGC...GGCGTGGCTACGG 6102
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
70 yValGlyValSerSerTyrglyTyrglyTyrrProSerTTPGlyTyrrProT 87
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
6103 CTCTGGAGGTAGCAGCTATGCTCCGAGGTGCTAGCTATGTTCTGGAG 6152
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
87 yrclyGlyTyrglyTyrglyTyrglyTyrglyTyrrProSerTTPGlyTyrrProT 103
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
6153 GTGGCGGGCGGGCGCGCGCTGCGAGCTATGCTCC...GGAGGTGGCAGC 6199
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
104 GlnGlyPheGlySerAlaTyrglyGlyTyrrProGlyTyrrTyrglyTyrrTy 120
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
6200 TATGGCTCTGGAGGTGGCGGGCGGCGCATGGCAGCTACCGCTCCGGAAG 6249
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
120 rTyrrProSerGlyTyrglyGlyTyrglyTyrglyTyrrTyrrTyrrTyrr 137
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
6250 CACACAGTGGGGCTACAGAGGTGCTCTGGAGGCGGGCGGCGGCGAGCT 6299
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
137 yrclyGlySerTyrrTyrrProAsnValArgAlaSerAlaGlyAlaAla 153
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
6300 CTGGCGGC.....CGGGCTCTGGCGGGGAGC 6328
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
154 Ala 154
6329 TCT 6331

```

seq_name: /cgn2_6/ptodata/2/1na/5B_COMB.seq: US-08-147-777-3

seq_documentation_block:

Sequence 3, Application US/08147777

Patent No. 5914265

GENERAL INFORMATION:

APPLICANT: Roop, Dennis R.
APPLICANT: Rothnagel, Joseph A.
APPLICANT: Greenhalgh, David A.
APPLICANT: Yuspa, Stuart H.
TITLE OF INVENTION: KERATIN K1 EXPRESSION VECTORS
TITLE OF INVENTION: AND METHODS OF USE
NUMBER OF SEQUENCES: 5
CORRESPONDENCE ADDRESS:
ADDRESSEE: LYON & LYON
STREET: 611 West Sixth Street
CITY: Los Angeles
STATE: California
COUNTRY: U.S.A.
ZIP: 90017

COMPUTER READABLE FORM:

MEDIUM TYPE: 3.5" diskette, 1.44 Mb storage

COMPUTER: IBM PC compatible

OPERATING SYSTEM: IBM MS-DOS (Version 5.0)

SOFTWARE: WordPerfect (Version 5.1)

[illegible]

LOCATION: 219..2830
US-09-010-928B-1

alignment_scores:
Quality: 154.50 Length: 138
Ratio: 2.207 Gaps: 7
Percent Similarity: 50.725 Percent Identity: 37.681

alignment_block:
US-09-554-547-16 x US-09-010-928B-1 ..

Align seg 1/1 to: US-09-010-928B-1 from: 1 to: 2830

```
46 GLYPProAGGLYtyr...ProGlyLeuThrAlaSerIleGly...GLYGI 60
||||| ||| ||||| ||||| ||||| ||||| |||||
1935 GGACCTGGTGGCTCCGGACCAAGAGGCGTAGGATCTGGTGGCTCCGACA 1984
60 uValGlyAlaArgLeuGlyGlyAArgAlaGlyValGlyValSerSerTyrc 77
: ||||| ||||| ||||| ||||| ||||| ||||| |||||
1985 AGGAGGAGTAAGACCTAGTGGCTCCGGACCAAGTGGCGTAGGACTGGAG 2034
77 LYTYrGlyTYrProSerTrpGlyTYrProTyrcLy..... 88
|| ||| ||| ||||| ||||| ||||| |||||
2035 GCGTAGGACCGCGTGGTCTGGAGACCTTAGCGTCTGGTGGTCCGGA 2084
89 .....GlyTYrGlyGlyTYrGI 94
2085 CCCGGAAGTGCAGGAGCGCTGGAGACTTATGACCTGGTGGTTTCGG 2134
94 YGlyTYrGlyGlyTYrGlyTYrAspGlnGlyPheGlySerAlaTyrc 111
||| ||||| ||||| ||||| ||||| ||||| |||||
2135 AGGACCGCGTGGTTTCGAGGA...CCGCGTGGTGGTGGAGACCTAGC 2181
111 LY.....GlyTYrProGlyTYrTYrGlyTYr 119
|| ||| ||| ||||| ||||| ||||| |||||
2182 GTCCAGGTGGTCTGGTGGACCTAGCGACCAAGTGGTGGTGGAGCCC 2231
120 TYrTYrProSerGlyTYrGlyGlyTYr.....GlyGI 131
|| ||||| ||||| ||||| ||||| ||||| |||||
2232 TACGGACCAAGTGGTGGTGGACCTAGCGGCGCGGTGGTGGTGG 2281
131 YSerTYrGly...GlySerTYrGlyGlySerTYrThrTYrProAsnValA 147
| ||||| ||||| ||||| ||||| ||||| |||||
2282 ACCCTACGGGCGCGGAGGTGCTGGTGGATCTAGCGGCTGGGTGGTGG 2331
147 rGAlaSerAlaGly 151
::: ||||| |||||
2332 GTGGATCAGAGAGCT 2345
```

THIS PAGE BLANK (USPTO)

Tel: 314 286 1800
 Fax: 314 286 1810
 Email: mouseest@watson.wustl.edu
 This clone is available royalty-free through LNL; contact the
 IMAGE Consortium (info@image.llnl.gov) for further information.
 MGI:672975

Seq primer: -28m13 rev1 ET from Amersham
 High quality sequence stop: 446.

FEATURES

source

1.494
 /organism="Mus musculus"
 /strain="C57BL/6"
 /db_xref="taxon:10090"
 /clone_image="1281175"
 /clone_lib="Stratagene mouse skin (#937313)"
 /sex="females"
 /tissue_type="whole skin"
 /dev_stage="11 weeks old"
 /lab_host="SOLR (kanamycin resistant)"
 /note="Organ: skin; Vector: pBluescript SK-; Site_1: EcoRI
 ; Site_2: XhoI; Cloned unidirectionally. Primer: Oligo
 dt. Whole skin from 11 week old C57BL/6 female mice.
 Average insert size: 1.0 kb; Uni-ZAP XR Vector; -5'
 adaptor sequence: 5' GAATTCGGCAGCAG 3' -3' adaptor
 sequence: 5' CTCGAGCTTTT TTTT TTTT 3'."
 BASE COUNT 87 a 102 c 165 g 140 t
 ORIGIN

alignment_scores:

Quality: 208.50 Length: 112
 Ratio: 2.780 Gaps: 10
 Percent Similarity: 66.964 Percent Identity: 47.321

alignment_block:

US-09-554-547-16 x AA799268 ..

Align seg 1/1 to: AA799268 from: 1 to: 494

```

49 G1yTyPrG1yLeuThra1aSer1leG1yG1uValG1yAla..... 63
||||| ||| :::::||||| :::::
34 GGCCTAC...GGCTATGCCAGCAGCTTTGGAGGC...CTGGCTGTGGAG 77
64 .....ATGLeuG1yG1yArG1aG1yAlaG1yAlaSer1 76
||||| ||| :::::|||||
78 TAACAGCATCCGACACTGGCTGTGGCTGTGGCTATGAGGCTTCGCAT 127
76 yrg1yTyrg1yTyPrProSerTrp...G1yTyPrProTyrg1y...G1yTyPr 90
||||| ||| :::::|||||
128 ATGCGCTGTGGCTATGAGGCTTTGGAGGCTTTGGATATGCGCTGTGGCTAT 177
91 G1yG1yTyrg1y.....G1yTyrg1yG1yTyrg1yTyrg1yTAspG1 104
||||| ||| :::::|||||
178 GGAGGCTATGAGGATACGCTGTGGCTATGAGGCTTTGGAGGCTTT..... 222
104 ng1yPhg1ySer1aTyrg1y.....G1yTyPrProG 115
||||| ||| :::::|||||
223 GGATATGGCTGTGGCTATGAGGCTTCGCATATGCGCTGTGGCTATGAG 271
115 1yTyPrTyrg1yTyTyTyPrProSerG1yTyrg1yG1y...G1yTyrg1y 130
||||| ||| :::::|||||
272 GCTTTGGAGGCTTTGGATATGCGCTGTGGCTATGAGGCTATGAGATATGCG 321
131 G1ySerTyrg1yG1y...SerTyrg1yG1ySerTyrg1y 141
:::::||||| :::::|||||
322 TCTGCGCTTTGGAGGCTATGAGATATGCGCTGTGGCTTC 357

```

seq_name: gb_est1:AA792705

seq_documentation_block: 578 bp mRNA EST 09-FEB-1998
 LOCUS AA792705
 DEFINITION vs75d11.r1 StrataGene mouse skin (#937313) Mus musculus cDNA clone
 IMAGE:1152117 5', mRNA sequence.

ACCESSION AA792705
 VERSION AA792705.1 GI:2855660
 KEYWORDS EST.
 SOURCE house mouse.
 ORGANISM Mus musculus.

REFERENCE

1 (bases 1 to 578)
 Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
 Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
 Teisling,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
 Waterston,R.

TITLE

The WashU-HMNI Mouse EST Project
 Unpublished (1996)

COMMENT

Contact: Marra M/Mouse EST Project
 WashU-HMNI Mouse EST Project
 Washington University School of Medicine
 444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
 Tel: 314 286 1800
 Fax: 314 286 1810

Email: mouseest@watson.wustl.edu
 This clone is available royalty-free through LNL; contact the
 IMAGE Consortium (info@image.llnl.gov) for further information.
 MGI:625325

Putative full length read
 vector to vector length is 619
 Seq primer: -28m13 rev1 ET from Amersham
 High quality sequence stop: 439.

FEATURES

source

1.578
 /organism="Mus musculus"
 /strain="C57BL/6"
 /db_xref="taxon:10090"
 /clone_image="1152117"
 /clone_lib="Stratagene mouse skin (#937313)"
 /sex="females"
 /tissue_type="whole skin"
 /dev_stage="11 weeks old"
 /lab_host="SOLR (kanamycin resistant)"
 /note="Organ: skin; Vector: pBluescript SK-; Site_1: EcoRI
 ; Site_2: XhoI; Cloned unidirectionally. Primer: Oligo
 dt. Whole skin from 11 week old C57BL/6 female mice.
 Average insert size: 1.0 kb; Uni-ZAP XR Vector; -5'
 adaptor sequence: 5' GAATTCGGCAGCAG 3' -3' adaptor
 sequence: 5' CTCGAGCTTTT TTTT TTTT 3'."
 BASE COUNT 115 a 119 c 178 g 166 t
 ORIGIN

alignment_scores:

Quality: 208.50 Length: 112
 Ratio: 2.780 Gaps: 10
 Percent Similarity: 66.964 Percent Identity: 47.321

alignment_block:

US-09-554-547-16 x AA792705 ..

Align seg 1/1 to: AA792705 from: 1 to: 578

```

49 G1yTyPrG1yLeuThra1aSer1leG1yG1uValG1yAla..... 63
||||| ||| :::::||||| :::::
22 GGCCTAC...GGCTATGCCAGCAGCTTTGGAGGC...CTGGCTGTGGAG 65
64 .....ATGLeuG1yG1yArG1aG1yAlaG1yAlaSer1 76
||||| ||| :::::|||||
66 TAACACATCCGACACTGGCTGTGGCTGTGGCTATGAGGCTTCGCAT 115
76 yrg1yTyrg1yTyPrProSerTrp...G1yTyPrProTyrg1y...G1yTyPr 90
||||| ||| :::::|||||
116 ATGCGCTGTGGCTATGAGGCTTTGGAGGCTTTGGATATGCGCTGTGGCTAT 165
91 G1yG1yTyrg1y.....G1yTyrg1yG1yTyrg1yTyrg1yTAspG1 104

```

```

|||||
166 GGAGGCTACGATACGCTCTGCTATGAGGCTTGGAGGCTTT..... 210
104 nglypneglyseralatyrgly.....glytyrpro 115
211 .GGATATGGCTCTGGCTATGAGGCTTCGATATGCTCGCTATGAG 259
115 lYTYrTYrTYrTYrTYrTYrProSerGlyTYrGlyGly...GlyTYrGly 130
260 GCTTGGAGGCTTGGATATGCTCTGGCTATGAGGCTATGATATGCG 309
131 GlySerTYrGlyGly...SerTYrGlyGlySerTYr 141
310 TCTGGCTTGGAGGCTATGATATGCTCTGGCTTC 345

```

seq_name: gb_est1:AA500430

seq_documentation_block:

LOCUS AA500430 449 bp mRNA EST 01-JUL-1997
DEFINITION v181d03.r1 Scratogene mouse skin (#937313) Mus musculus cDNA clone
IMAGE:918629 5', mRNA sequence.

ACCESSION AA500430

VERSION AA500430.1 GI:2235397

KEYWORDS EST.

SOURCE house mouse.

ORGANISM Mus musculus

REFERENCE 1 (bases 1 to 449)

AUTHORS Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,

Gelsel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,

Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,

Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and

Waterston,R.

Unpublished (1996)

Contact: Marra M/Mouse EST Project

Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108

Tel: 314 286 1800

Fax: 314 286 1810

Email: mouseest@watson.wustl.edu

This clone is available royalty-free through LNL: contact the

IMGE Consortium (info@image.llnl.gov) for further information.

MGI:530845

Seq primer: -28m13 rev1 ET from Amersham

High quality sequence stop: 408.

Location/Qualifiers

1..449

/organism="Mus musculus"

/strain="C57BL/6"

/db_xref="taxon:10090"

/clone="IMAGE:918629"

/tissue="females"

/sex="females"

/dev_stage="11 weeks old"

/lab_host="SOLR (kanamycin resistant)"

/note="Organ: skin; Vector: pBluescript SK-; Site:1: EORI

; Site:2: XhoI; Cloned unidirectionally. Primer: Oligo

dt. Whole skin from 11 week old C57BL/6 female mice.

Average insert size: 1.0 kb; Uni-ZAP XR Vector; -5'

adaptor sequence: 5' GAATTCGGCAGCAG 3' -3' adaptor

sequence: 5' CTCGAGTTTCTTTTCTTTTCTTTT 3'

BASE COUNT 78 a 88 c 150 g 133 t

ORIGIN

alignment_scores:

Quality: 203.00 Length: 106

Ratio: 2.819 Gaps: 9

Percent Similarity: 67.925 Percent Identity: 47.170

alignment_block:

us-09-554-547-16 x AA500430 ..

Align seg 1/1 to: AA500430 from: 1 to: 449

```

55 AlAsertlEgLyGlyAluValGlyAla.....ArgLe 65
8 AGCAGCTTGGAGGC...CTGGGCTGTGATGTAAACAGCATCCCGAGCT 54
65 uGlyGlyArgAlaGlyAluValSerSerTYrGlyTYrTYrPro 82
55 GGGCTGTGGCTCTGGCTATGAGGCTTCGGATATGAGCTCTGGCTATGAG 104
82 eTRP...GlyTYrProTYrGly...GlyTYrGlyGlyTYrGly..... 94
105 GCTTGGAGGCTTGGATATGCTCTGGCTATGAGGCTACGATACGGC 154
95 ..GlyTYrGlyGlyTYrGlyGlyTYrAspGlnGlyPheGlySerAlaTy 110
155 TCTGGCTATGAGGCTTGGAGGCTTT.....GGATATGCTCTGGCTA 198
110 rGly.....GlyTYrProGlyTYrTYrGlyTYrTYrT 121
199 TGGAGGCTTCGGATATGCTCTGGCTATGAGGCTTGGAGGCTTTGGAT 248
121 YrProSerGlyTYrGlyGly...GlyTYrGlyGlySerTYrGlyGly... 135
249 ATGCTCTGGCTATGAGGCTATGATATGCTCTGGCTTGGAGGCTAT 298
136 SerTYrGlyGlySerTYr 141
299 GGATATGCTCTGGCTTC 316

```

seq_name: gb_est1:AA727403

seq_documentation_block:

LOCUS AA727403 469 bp mRNA EST 02-JAN-1998
DEFINITION vu99f12.r1 Stratagene mouse skin (#937313) Mus musculus cDNA clone
IMAGE:1210319 5', mRNA sequence.

ACCESSION AA727403

VERSION AA727403.1 GI:2745110

KEYWORDS EST.

SOURCE house mouse.

ORGANISM Mus musculus

REFERENCE 1 (bases 1 to 469)

AUTHORS Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,

Gelsel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,

Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,

Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and

Waterston,R.

Unpublished (1996)

Contact: Marra M/Mouse EST Project

Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108

Tel: 314 286 1800

Fax: 314 286 1810

Email: mouseest@watson.wustl.edu

This clone is available royalty-free through LNL: contact the

IMGE Consortium (info@image.llnl.gov) for further information.

Seq primer: -28m13 rev1 ET from Amersham

High quality sequence stop: 441.

Location/Qualifiers

1..469

/organism="Mus musculus"

/strain="C57BL/6"

/db_xref="taxon:10090"

/clone="IMAGE:1210319"

```

/clone_11b="Stratagene mouse skin (#937313)"
/sex="females"
/tissue.type="whole skin"
/dev_stage="11 weeks old"
/lab_host="SOLR (kanamycin resistant)"
/note="Organ: skin; Vector: pBluescript SK-; Site_1: EcoRI
; Site_2: XhoI; Cloned unidirectionally. Primer: Oligo
dr. Whole skin from 11 week old C57BL/6 female mice.
Average insert size: 1.0 kb; Uni-ZAP XR Vector: -5'
adaptor sequence: 5' CATTCGCCACGAC 3' -3' adaptor
sequence: 5' CTCGAGTTTCTTTTCTTTTCTTTT 3'"

BASE COUNT      78 a      94 c      158 g      139 t
ORIGIN

alignment_scores:
  Quality: 203.00      Length: 113
  Ratio: 2.942      Gaps: 9
  Percent Similarity: 61.062      Percent Identity: 46.018

alignment_block:
US-09-554-547-16 x AA727403 ..

Align seg 1/1 to: AA727403 from: 1 to: 469

53 LeuThrAlaSerIleGlyGluValGlyAlaArgLeuGlyArgAl 69
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
13 CTACTCCGAGCGCTGGCTACGCGCTATGGCAGCCTTTGGAGGC...CT 59

69 aGlyValGlyValSerSerTyr.....GlyTyrGlyTyrP 81
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
60 GCGCTGTGATGTACAGCATCCCATCTGGCTGTGCTGCTGCTATG 109

81 roSerTrpGlyTyrProTyrGlyGlyTyrGlyTyrGlyTyrGly 97
   : : : : : : : : : : : : : : : : : : : : : : : :
110 TAGGCTTCGATATGCG...TCGTGCTATGGAGGCTTGGAGGCTTTGCA 156

98 .....GlyTyrGlyGlyTyrAspGlnGlyPheGlySerAlaTyrG 111
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
157 TATGGCTTCGCTATGGAGCGCTAC.....CGATACGCGCTCGCTATG 200

111 yGlyTyr.....ProGlyTyrTyrGlyTyrTyrTyrP 122
   ||| : : : : : ||| ||| ||| ||| ||| ||| ||| |||
201 AGGCTTTGGAGGCTTGGATATGCTGCTGATGAGGCTTCGGATATG 250

122 roSerGlyTyrGlyGly.....GlyTyrGlyGlySerTyrGly 134
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
251 GCTGTGCTATGGAGGCTTGGAGGCTTGGATATGCTGCTGCTATGGA 300

135 Gly.....SerTyrGly.....GlySerTyrThrTyr 143
   ||| : : : ||| ||| ||| ||| ||| ||| ||| ||| |||
301 GCCTATGATATGCTTCGCTTGGAGGCTATGATAT 339

seq_name: gb_est1:BE196524

seq_documentation_block:
LOCUS      BE196524      962 bp      mRNA      EST      02-MAR-2001
DEFINITION HVSMEH0092116f Hordeum vulgare 5-45 DAP spike EST library
            HVCNDA0009 (5 to 45 DAP) Hordeum vulgare cDNA clone HVSMEH0092116f,
            mRNA sequence.
ACCESSION  BE196524
VERSION    BE196524.1 GI:8708720
KEYWORDS   EST.
SOURCE     barley.
            Hordeum vulgare
            Eukaryota: Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidae
            ; Triticeae; Hordeum.
            1 (bases 1 to 962)
REFERENCE  1 (bases 1 to 962)
AUTHORS   Wing,R., Close,T.J., Kleinholz,A., Wise,R., Begum,D., Fritsch,D., Yu
            Y., Anderson,H., Dale,J., Henry,D., Kernodle,S., Palmer,M., Rambo
            T., Snakl,C., Schwartzbeck,J., Simmons,J., Chol,D.W., Main,D. and
            Wood,T.

```

```

TITLE      Development of a genetically and physically anchored EST resource
            for barley genomics
JOURNAL    Unpublished (2000)
COMMENT    Contact: Wing RA
            Clemson University Genomics Institute
            100 Jordan Hall, Clemson, SC 29634, USA
            Tel: 864 656 7288
            Fax: 864 656 4293
            Email: rwing@clemson.edu
            Seq primer: AATTACCTTCACCTAAAGG
            High quality sequence stop: 723.

FEATURES   Location/Qualifiers
source     1..962
            /organism="Hordeum vulgare"
            /cultivar="Morex"
            /db_xref="taxon:4513"
            /clone="HVSMEH0092116f"
            /clone_11b="Hordeum vulgare 5-45 DAP spike EST library
            HVCNDA0009 (5 to 45 DAP)"
            /tissue.type="5-45 DAP Spike"
            /lab_host="SOLR"
            /note="Vector: lambdaZAP; Site_1: EcoRI; Site_2: XhoI; For
            more details on library preparation and sequence analysis
            see http://www.genome.clemson.edu/projects/barley/ To
            order a clone see http://www.genome.clemson.edu/orders"

BASE COUNT      242 a      185 c      309 g      224 t      2 others
ORIGIN

alignment_scores:
  Quality: 202.50      Length: 162
  Ratio: 2.355      Gaps: 11
  Percent Similarity: 53.086      Percent Identity: 40.741

alignment_block:
US-09-554-547-16 x BE196524 ..

Align seg 1/1 to: BE196524 from: 1 to: 962

23 ArgLeuGlySerAspLeuAspThrPheGlyArgValHis.....GlyAs 37
   ||| : : : ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
86 AGGCTTATGTCTAGAT.....GGGAGATCATGATCTTGAGAG 123

37 nleuTyrAlaGlyIleGluArgAlaGlyProArg..... 48
   : : : : : : : : : : : : : : : : : : : : : : : :
124 GAAGCAGTTGAATMAAAGGCCGACCAAGGAAACCTGGTGGTG 173

49 .....GlyTyrProGlyLeuThrAla 55
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
174 ATTCAAGCTCTAACAGTACATAGCATGAGAGAGTGCCACCGTACT 223

56 SerIleGlyGlyGluValGlyAlaArgLeuGlyGlyArgAlaGlyVal 72
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
224 TCATACCGTGTGGAGGCGCGCGAGTGTGTGGAGCGCTTCATGATCAG 273

72 yAlaSerSer...TyrGlyTyrGly..... 79
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
274 CAGCAGCAGGCGCTATGCGATGAGCTGACTACCGATCAGTCGACAGC 323

80 .....TyrProSerTrpGlyTyrProTyrGlyGlyTyrGly.....GlyTyr 93
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
324 CCTACTATGATGTAGCAGCA.....TATGTGCTTACGCTTACAGGCTAT 367

94 GlyGlyTyrGlyGlyTyrGlyGlyTyrAspGlnGlyPheGlySerAlaTyr 110
   ||| : : : ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
368 GGTGCAATATGAGAGTAACTCCCTATGCGTGTGGAGGCTTGGCTGCTGTA 417

110 rGlyGly.....TyrProGlyTyrTyrGlyTyrTyrTyrProSerGlyT 125
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
418 TGGTGTCTTATATATGAGAGGCCGATATGCTGATAT..... 454

125 yGlyGlyGlyTyrGlyGlySerTyrGlyGlySerTyrGlyGlySerTyr 141

```

455GGGCGATATGGGGTGGCTATGA.....GGTGGCGCATAT 490
 142 ThrTyPro...AsnValArgAlaSerAlaGlyAla 152
 491 GGTACACCTGTGGCTATGTGTGAGGTGACGTGCA 526

seq_name: gb_est2:BG403710

seq_documentation_block:
 LOCUS BG403710 766 bp mRNA EST 12-MAR-2001
 DEFINITION 602419538P1 NIH_MGC_93 Homo sapiens cDNA clone IMAGE:4526721 5',
 mRNA sequence.

ACCESSION BG403710 GI:13297158

VERSION BG403710.1

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
 1 (bases 1 to 766)
 NIH-MGC http://mgc.nci.nih.gov/.

AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)

TITLE Unpublished (1999)

COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgapds-remail.nih.gov
 Tissue Procurement: ATCC

CDNA Library Preparation: Life Technologies, Inc.
 DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
 http://image.llnl.gov

Plate: LHAM10434 row: k column: 10

High quality sequence stop: 729.

Location/Qualifiers

FEATURES

SOURCE

1..766

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="IMAGE:4526721"

/clone_lib="NIH-MGC_93"

/lab_host="DH10B (phage-resistant)"

/note="Organ: bladder; Vector: pCMV-SPORT6; Site_1: NotI;
 Site_2: SalI; Cloned unidirectionally; oligo-dr primed.
 Average insert size 1.7 kb. Library enriched for
 full-length clones and constructed by Life Technologies.
 Note: this is a NIH-MGC Library."

BASE COUNT 184 a 109 c 338 g 135 t

ORIGIN

alignment_scores:

Quality: 201.50 Length: 142
 Ratio: 2.457 Gaps: 9
 Percent Similarity: 57.746 Percent Identity: 42.958

alignment_block:

US-09-554-547-16 x BG403710 ..

Align seg 1/1 to: BG403710 from: 1 to: 766

20 ArgAlaGlyArgLeuGlySerAspLeuAspThrPheGlyArgValHisG1 36

242 AGAGCGCGCGCTATGTGTGGGCGAC.....AGAGG 270

36 yasnleuryralaglylliegluargalaglyProargGlyTyProGlyU 53

271 AGCGCGCTATGGAGG...GACCGAGGAGGT.....GGCTATGAGGAG 311

53 eutThalaserllieglygluvalaGlyAlaArgLeuGlyGly...Arg 68

312 ATCGAGGTGCTATGGAGGAGACCGAGGTGAGGCTATGTGTGAGACCGA 361

69 AlaglyValaGlySerSerTyrglyTyrglyTyrglyProSerTyrglyTy 85
 362 GGAGGCTATGAGAGATCGAGAGGTTACGAGAGATCGAGAGAGGTTA 411
 85 r.....ProtyrglyTyrglyTyrglyTyrglyTyrglyTyrglyTyrg 100
 412 TGGAGGAGATCGAGAGGCTATGGAGGAGACAGAGCGGGGGGCTATG 461
 100 LysGlyTyrglyTyrglyTyrglyTyrglyTyrglyTyrglyTyrglyTy 114
 462 GAGGA...GACCGTGTGTGTGGCTATGCGTATGAGAGACCGAGTGA 508
 115 GlyTyrglyTyrglyTyrglyTyrglyTyrglyTyrglyTyrglyTyrg 127
 509 GGCTATGAGAGAGACAGAGATGTGTGGCTATGAGAGACCGAGTGTG 558
 127 yGlyTyrglyTyrglyTyrglyTyrglyTyrglyTyrglyTyrglyTyrg 141
 559 GGGCTACGAGAGACAGACCGAGTGTGTATGAGGCAAAATGGAGAGAA 608
 141 yTrhTyrglyProAsnValArgAlaSer 149
 609 ACCACTACAGAAATGATCAGGCAACG 634

seq_name: gb_est1:AL372330

seq_documentation_block: 435 bp mRNA EST 03-AUG-2000

LOCUS AL372330 MTBA50C01F1 MTBA Medicago truncatula cDNA clone MTBA50C01 T3, mRNA

DEFINITION MTBA50C01F1 MTBA Medicago truncatula cDNA clone MTBA50C01 T3, mRNA

ACCESSION AL372330

VERSION AL372330.1 GI:9672083

KEYWORDS EST.

SOURCE barrel medic.

ORGANISM Medicago truncatula

REFERENCE Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae;
 Medicago.
 1 (bases 1 to 435)

AUTHORS Journeé,E.P., Crespeau,H., van-Tuluen,D., Gouzy,J., Jallion,O.,
 Niebel,A., Carreau,V., Chataigner,O., Kahn,D., GlaninaZZI-Pearson
 'V. and Gamas,P.

Medicago truncatula ESTs from nitrogen-starved roots

Unpublished (2000)

CONTACT: Genoscope

Genoscope - Centre National de Sequencage

BP 191 91006 Evry cedex - France

Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr

Contact : Pascal Gamas and Etienne-Pascal Journeé, Laboratoire de

Biologie Moléculaire des Relations Plantes-Microorganismes,

CNRS-INRA, BP 27 31326 Castanet-Tolosan Cedex, France (Email :

mt-est@genoscope.inra.fr Website :

http://sequence.coulouze.inra.fr/truncatula.html).

Location/Qualifiers

1..435

/organism="Medicago truncatula"

/cultivar="Jemalong"

/db_xref="taxon:3880"

/clone="MTBA50C01"

/tissue_type="root tips"

/dev_stage="harvested after 3 days of N-starvation"

/note="Vector: pBluescript PSK; Site1: EcoRI; Site 2:
 XhoI. Plants were grown in an aeroponic chamber for 14
 days on nitrogen-rich medium followed by 3 days on N-free
 medium. RNA was extracted from root tips (1-3 cm). cDNA
 was prepared from polyA+ enriched RNA. The cDNA was
 directionally ligated into Uni-zapR vector from
 stragene and packaged using GibcoBRL Gold packaging
 extracts. Plasmids containing cDNA inserts were
 mass-excised from phage stocks using Exasit helper phage

KEYWORDS EST

```

SOURCE          zebrafish.
ORGANISM        Danio rerio.
REFERENCE       Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS         Actinopterygii; Neopterygii; Teleostei; Euteleostei; Ostariophysi;
                Cypriniformes; Cyprinidae; Rasbora; Danio.
TITLE           1 (bases 1 to 504)
JOURNAL         Clark,M., Johnson,S.L., Lehrach,H., Lee,R., Li,F., Marra,M., Eddy
COMMENT         S., Hillier,L., Kucab,T., Martin,J., Beck,C., Wyllie,T., Underwood
                R., Steptoe,M., Theising,B., Allen,M., Bowers,Y., Person,B.,
                Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk,R., Ritter,E.,
                Kohn,S., Shih,T., Jackson,Y., Cardenas,M., McCann,R., Waterson,R.,
                and Wilson,R.
                Mashu zebrafish EST Project 1998
                Unpublished (1998)
                Contact: Stephen L. Johnson
                Washington University School of Medicine
                444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
                Tel: 314 286 1800
                Fax: 314 286 1810
                Email: zbrafish@watson.wustl.edu
                cDNA library preparation: John Ngai cDNA library Arrayed by:
                Matthew Clark. DNA Sequencing by: Washington University Genome
                Sequencing Center Clone Distribution: Genome Systems, St. Louis,
                Missouri (web address: www.genomesystems.com) (email contact:
                info@genomesys.wustl.edu) and Research Genetics, Huntsville, Alabama
                (web address: www.resgen.com) (email contact: info@resgen.com) and
                ResourcenzentrumPrimatendank, Berlin, Germany (web address:
                www.rtpd.de)
                Seq primer: T3 ET from Amersham
                High quality sequence stop: A40.
FEATURES
source          Location/Qualifiers
                1..504
                /organism="Danio rerio"
                /db_xref="taxon:7955"
                /clone_id="zebrafish adult olfactory"
                /sex="mixed"
                /issue_type="Olfactory rosettes"
                /ded_stage="adult"
                /lab_host="D10Hb (Gibco BRU)"
                /note="Vector: pSPOR1; Site_1: NotI; Site_2: SalI; This
                is a directionally cloned cDNA library from adult
                zebrafish olfactory epithelium."
BASE COUNT      104 a      120 c      200 g      80 t
ORIGIN
alignment_scores:
Quality: 198.00 Length: 100
Ratio: 3.094 Gaps: 6
Percent Similarity: 64.000 Percent Identity: 48.000
alignment_block:
US-09-554-547-16 x AW233247 ..
Align seg 1/1 to: AM233247 from: 1 to: 504
58 GlyGlyGlyValGlyAlaArg...LeuGlyGlyATGAGlaGlyValGlyA 73
||||| ||||:|||| | ||||| ||||:||||
37 GTGTGGCGCCGCCGCGGCAGAGGAGAGCGGAGAGAAGCATGGCGCG 86
||||| ||||:|||| | ||||| ||||:||||
73 lserseTyTgLyTYrTYrProseTyrrpGlyTYrroTYrGlyGlyT 90
||||| ||||:|||| | ||||| ||||:||||
87 CAATCAAAATGGCTTCGGC.....GGCGTAGAGAGAGGCGCT 124
||||| ||||:|||| | ||||| ||||:||||
90 yrgLyGlyTYrGlyGly.....TYrGlyGlyTYrGlyGlyTYrAspGln 104
||||| ||||:|||| | ||||| ||||:||||
125 AGCGCGCGCTACGGCGCGCTATGGCGGAGATACGGCGGAACAATGCGC 174
||||| ||||:|||| | ||||| ||||:||||
105 glyPheglySerAlaTyTyrGlyTYrProGlyTYrTYrTYrTYrTY 121
||||| ||||:|||| | ||||| ||||:||||
175 GSCTACGGCGCTACGATGTAAGT.....GGAGCGTACGCT..... 210
||||| ||||:|||| | ||||| ||||:||||
121 rProseTyTyrGlyGlyGlyTYrGlyGlySerTYrGly.....G 135

```

```

211 ... GGGCGTTACGAGGATGCGATACGAGAGGACTTACGTCGACCAATGG 256
135 lYSeTyrIGlYglSeTyrThrTYrProAsnValAlaArgAlaSerAlaGly 151
|||||
|||||
GGGGCGATATGCGCGATGGAGACGGTTACAGAGCTTGGCGACGGATACGCGC 306
257

seq_name: gb_est1:A1878240

seq_documentation_block:
LOCUS      A1878240      531 bp      mRNA      EST      07-JUN-2001
DEFINITION zebrafish Washu M1NG EST Danio rerio cDNA clone
IMAGE:3725012 5' similar to SW:R031.XENLA P51968 HETEROGENEOUS
NUCLEAR RIBONUCLEOPROTEIN A3 HOMOLOG 1: contains element MRR22
repetitive element ; mRNA sequence.
A1878240
A1878240.1 GI:5552370
EST.
zebrafish.
ORGANISM   Danio rerio
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Ostariophysi;
Cypriniformes; Cyprinidae; Rasbora; Danio.
REFERENCE 1 (bases 1 to 531)
AUTHORS   Clark,M., Johnson,S.L., Lehrach,H., Lee,R., Li,F., Marra,M., Eddy
,S., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T., Underwood
,K., Steptoe,M., Theising,B., Allen,M., Bowers,Y., Person,B.,
Swaller,T., Gibbons,M., Page,D., Harvey,N., Schurk,R., Ritter,E.,
Kohn,S., Shih,T., Jackson,Y., Cardenas,M., McCann,R., Waterston,R.,
and Wilson,R.
Washu zebrafish EST Project 1998
Unpublished (1998)
Other-ESTs: fc52c11.x1
Contact: Stephen L. Johnson
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: zdratish@watson.wustl.edu
cDNA Library Preparation: Matthew Clark, cDNA Library Arrayed by:
Matthew Clark, DNA Sequencing by: Washington University Genome
Sequencing Center Clone Distribution: Genome Systems, St. Louis,
Missouri (web address: www.genomesystems.com) (email contact:
info@genomesystems.com) and Research Genetics, Huntsville, Alabama
(web address: www.resgen.com) (email contact: info@resgen.com) and
ResourcenCenter@umprimaridenbank, Berlin, Germany (web address:
www.rzp.d.de)
Seq primer: T3 ET from Amersham
High quality sequence, stop: 477.
Location/Qualifiers
1..531
/organism="Danio rerio"
/db_xref="taxon:7955"
/clone="IMAGE:3725012"
/clone_ltb="zebrafish Washu M1NG EST"
/sex="mixed"
/tissue_type="26 somite embryos, adult livers, shield
stage embryos"
/lab_host="XLI-blue MRF"
/notes="vector: pSPORT1; site_1: NotI; site_2: SalI; 1st
strand cDNA was primed with a Not I... oligo(dT)15 primer
[5' pGACTGCTTCAGATCGCGAGCGCGCCCTTTTCTTTTCTTTT3' ];
double-stranded cDNA was ligated to Sal I adaptors (BRL),
digested with Not I and cloned into the Not I and Sal I
sites of the pSPORT1 vector (BRL). Library was constructed
by Matthew Clark (Lehrach lab, ICRF, London and Max Planck
Institut fuer Molekulare Genetik, Berlin). cDNAs for EST
analyses were selected following oligonucleotide
hybridization fingerprinting of arrayed clones from
zebrafish late somitogenesis (26 ss), adult liver or
embryonic shield stage (5.6 h) libraries. Fingerprint
data were used to computationally cluster cDNAs, and a
single cDNA from each cluster was chosen for sequencing.

```



```

seq_documentation_block:
LOCUS      A1795569      629 bp      mRNA      EST      02-JUL-1999
DEFINITION 614009H04.x2 614 - root cDNA library from Walbot Lab Zea mays cDNA,
            mRNA sequence.
ACCESSION  A1795569
VERSION    A1795569.1 GI:5343264
KEYWORDS   EST.
SOURCE     Zea mays.
ORGANISM   Zea mays.
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
            clade; Panicoideae; Andropogoneae; Zea.
REFERENCE  1 (bases 1 to 629)
AUTHORS   Walbot,V.
TITLE     Maize ESTs from various cDNA libraries sequenced at Stanford
            University
JOURNAL    Unpublished (1999)
COMMENT    Contact: Walbot V
            Department of Biological Sciences
            Stanford University
            855 California Ave, Palo Alto, CA 94304, USA
            Tel: 650 723 2227
            Fax: 650 725 8221
            Email: walbot@stanford.edu
            Plate: 614009 row: H column: 04.
            Location/Qualifiers
                source          1..629
                /organism="Zea mays"
                /cultivar="W23"
                /db_xref="taxon:4577"
                /clone_id="614 - root cDNA library from Walbot Lab"
                /tissue_type="root"
                /dev_stage="3-4 days old"
                /lab_host="XL0LR"
                /note="Organ: root; Vector: pBluescriptII SK+; Site_1:
                Ecort; Site_2: XhoI; 3-4 days old root tissue from Walbot
                Lab (LM)"
BASE COUNT 126 a 112 c 214 g 177 t
ORIGIN
alignment_scores:
    Quality: 191.00      Length: 119
    Ratio: 2.768        Gaps: 8
    Percent Similarity: 57.983      Percent Identity: 47.059
alignment_block:
US-09-554-547-16 x A1795569 ..
Align seg 1/1 to: A1795569 from: 1 to: 629
39  TTTAAGLIIleGLuARgAlaGLyPRoARgLYTyr.....ProGI 52
    |||::||| |||::||| |||::||| |||::|||
29  TATGCTGCTGCTTATGATCGATCTGTCGACGAGCTTATGATGATGATGCTG 78
    |||::||| |||::||| |||::||| |||::|||
52  yLeuThraIaSerIIeGLyGLyUaGLyAlaARgLeuGLyGLyARgA 69
    |||::||| |||::||| |||::||| |||::|||
79  CCCAGAGCAGCATATGCTATGCTATGCTATGCTATGCTATGCTATGCTATG 128
    |||::||| |||::||| |||::||| |||::|||
69  laGLyAlaGLyAlaSerSerTYrGLyTYrGLyTYrProSer..... 82
    |||::||| |||::||| |||::||| |||::|||
129  CTGGCTTGGG...TCTACTTTTGGTGGTGGTGGTGGTGGTGGTGGTGGT 175
    |||::||| |||::||| |||::||| |||::|||
83  TTPGLYTYrPRoTYrGLyTYrGLyGLyTYrGLyGLy.....TYrGI 97
    |||::||| |||::||| |||::||| |||::|||
176  GGAGGTGCTGCTATGCTATGCTATGCTATGCTATGCTATGCTATGCTATG 225
    |||::||| |||::||| |||::||| |||::|||
97  yGLY.....TYrGLyGLyTYrAspGLInGLyPheGLySerAlaTYrGLyG 112
    |||::||| |||::||| |||::||| |||::|||
226  AGGAGGTGCTACGAGGCGGTGCTATGAGGAGGTGCTGCTGCTATGCTATG 275
    |||::||| |||::||| |||::||| |||::|||
112  lYTYrProGLyTYrGLyTYrTYrTYrProSerGLyTYrGLy..... 126

```

```

seq_name: gb_est1:AW000266
seq_documentation_block:
LOCUS      AW000266      649 bp      mRNA      EST      08-SEP-1999
DEFINITION 614009H04.Y1 614 - root cDNA library from Walbot Lab Zea mays cDNA,
            mRNA sequence.
ACCESSION  AW000266
VERSION    AW000266.1 GI:5847187
KEYWORDS   EST.
SOURCE     Zea mays.
ORGANISM   Zea mays.
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
            clade; Panicoideae; Andropogoneae; Zea.
REFERENCE  1 (bases 1 to 649)
AUTHORS   Walbot,V.
TITLE     Maize ESTs from various cDNA libraries sequenced at Stanford
            University
JOURNAL    Unpublished (1999)
COMMENT    Contact: Walbot V
            Department of Biological Sciences
            Stanford University
            855 California Ave, Palo Alto, CA 94304, USA
            Tel: 650 723 2227
            Fax: 650 725 8221
            Email: walbot@stanford.edu
            Plate: 614009 row: H column: 04.
            Location/Qualifiers
                source          1..649
                /organism="Zea mays"
                /cultivar="W23"
                /db_xref="taxon:4577"
                /clone_id="614 - root cDNA library from Walbot Lab"
                /tissue_type="root"
                /dev_stage="3-4 days old"
                /lab_host="XL0LR"
                /note="Organ: root; Vector: pBluescriptII SK+; Site_1:
                Ecort; Site_2: XhoI; 3-4 days old root tissue from Walbot
                Lab (LM)"
BASE COUNT 131 a 115 c 220 g 182 t 1 others
ORIGIN
alignment_scores:
    Quality: 191.00      Length: 119
    Ratio: 2.768        Gaps: 8
    Percent Similarity: 57.983      Percent Identity: 47.059
alignment_block:
US-09-554-547-16 x AW000266 ..
Align seg 1/1 to: AW000266 from: 1 to: 649
39  TTTAAGLIIleGLuARgAlaGLyPRoARgLYTyr.....ProGI 52
    |||::||| |||::||| |||::||| |||::|||
40  TATGCTGCTGCTTATGATCGATCTGTCGACGAGCTTATGATGATGATGCTG 89
    |||::||| |||::||| |||::||| |||::|||
52  yLeuThraIaSerIIeGLyGLyUaGLyAlaARgLeuGLyGLyARgA 69
    |||::||| |||::||| |||::||| |||::|||
90  CCAGAGCAGCATATGCTATGCTATGCTATGCTATGCTATGCTATGCTATG 139
    |||::||| |||::||| |||::||| |||::|||
69  laGLyAlaGLyAlaSerSerTYrGLyTYrGLyTYrProSer..... 82

```

```
140 CTGGCTTTGGG...TCTAGTTTGGTGGCTTATGCTGGATCCATGTAAT 186
83 TrpGlyTyrProTyrGlyGlyTyrGlyTyrGlyGly.....TyrGly 97
187 GGAGCTGCTGCCTATGCTGCATATGCTGCTATGCTGCTGCGCCTATGG 236
97 yGly.....TyrGlyGlyTyrAspGlnGlyPheGlySerAlaTyrGlyG 112
237 AGGAGGTGCCTACGAGGCGGCTGCTATGAGAGTGTGCTGCTGCTATGGTG 286
112 LyTyrProGlyTyrTyrGlyTyrTyrTyrProSerGlyTyrGly..... 126
287 CCCCCCAGCTGCTTATGGC.....ACTGGCGGATACGGCAGTTAT 327
127 GlyGlyTyrGlyGlySerTyrGlyGlySerTyrGly.....GlySerTyr 141
328 GCGGAGCAGGTAGTACTGCTGTGGGAGCACAGGTGCTCGGGGTTCTAG 377
141 rThrTyr 143
378 CAGGTAT 384
```

THIS PAGE BLANK (USPTO)

**This Page is Inserted by IFW Indexing and Scanning
Operations and is not part of the Official Record**

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images include but are not limited to the items checked:

- ☒ **BLACK BORDERS**
- ☐ **IMAGE CUT OFF AT TOP, BOTTOM OR SIDES**
- ☒ **FADED TEXT OR DRAWING**
- ☒ **BLURRED OR ILLEGIBLE TEXT OR DRAWING**
- ☐ **SKEWED/SLANTED IMAGES**
- ☐ **COLOR OR BLACK AND WHITE PHOTOGRAPHS**
- ☐ **GRAY SCALE DOCUMENTS**
- ☐ **LINES OR MARKS ON ORIGINAL DOCUMENT**
- ☐ **REFERENCE(S) OR EXHIBIT(S) SUBMITTED ARE POOR QUALITY**
- ☐ **OTHER:** _____

IMAGES ARE BEST AVAILABLE COPY.

As rescanning these documents will not correct the image problems checked, please do not report these problems to the IFW Image Problem Mailbox.

THIS PAGE BLANK (USPTO)